

THE 10TH JUBILEE CONFERENCE OF PHD STUDENTS IN COMPUTER SCIENCE

Volume of extended abstracts

CS^2

Organized by the Institute of Informatics of the University of Szeged



June 27 – June 29, 2016
Szeged, Hungary

Scientific Committee:

Zoltán Fülöp (Co-Chair, SZTE)
Lajos Rónyai (Co-Chair, SZTAKI, BME)
Péter Baranyi (SZE)
András Benczúr (ELTE)
Hassan Charaf (BME)
Tibor Csendes (SZTE)
László Cser (BCE)
Erzsébet Csuha-Varjú (ELTE)
János Demetrovics (ELTE)
Imre Dobos (BCE)
István Fazekas (DE)
† János Fodor (ÓE)
Aurél Galántai (ÓE)
Zoltán Gingl (SZTE)
Tibor Gyimóthy (SZTE)
Katalin Hangos (PE)
Zoltán Horváth (ELTE)
Csanád Imreh (SZTE)
Márk Jelasity (SZTE)
Zoltán Kása (Sapientia EMTE)
László Kóczy (SZE)
János Levendovszki (BME)
László Nyúl (SZTE)
Attila Pethő (DE)
Gábor Prószéki (PPKE)
Tamás Szirányi (SZTAKI)
Péter Szolgay (PPKE)
János Sztrik (DE)
János Tapolcai (BME)
János Végh (ME)

Organizing Committee:

Rudolf Ferenc, Balázs Bánhelyi, Tamás Gergely, Attila Kertész, Zoltán Kincses

Address of the Organizing Committee

c/o. Rudolf Ferenc
University of Szeged, Institute of Informatics
H-6701 Szeged, P.O. Box 652, Hungary
Phone: +36 62 546 396, Fax: +36 62 546 397
E-mail: cscs@inf.u-szeged.hu
URL: <http://www.inf.u-szeged.hu/~cscs/>

Sponsors

University of Szeged, Institute of Informatics
Evopro Group
Polygon Publisher

Preface

This conference is the tenth in a series. The organizers aimed to bring together PhD students working on any field of computer science and its applications to help them publishing one of their first abstracts and papers, and provide an opportunity to hold a scientific talk. As far as we know, this is one of the few such conferences. The aims of the scientific meeting were determined on the council meeting of the Hungarian PhD Schools in Informatics: it should

- provide a forum for PhD students in computer science to discuss their ideas and research results;
- give a possibility to have constructive criticism before they present the results at professional conferences;
- promote the publication of their results in the form of fully refereed journal articles; and finally,
- promote hopefully fruitful research collaboration among the participants.

The papers emerging from the presented talks will be forwarded to the Acta Cybernetica journal.

The organizers hope that the conference will be a valuable contribution to the research of the participants, and wish a pleasant stay in Szeged.

Szeged, June 2016

*Rudolf Ferenc
Balázs Bánhelyi
Tamás Gergely
Attila Kertész
Zoltán Kincses*

Contents

Preface	i
Contents	ii
Program	iv
Plenary talks	1
Dániel Marx: <i>The Optimality Program in Parameterized Algorithms</i>	1
László Pyber: <i>How to avoid the Classification Theorem of Finite Simple Groups in Asymptotic Group Theory</i>	2
Dániel Varró: <i>Models and Queries for Smart and Safe Cyber-physical Systems</i>	3
Abstracts	4
Bence Babati, Norbert Pataki, Zoltán Porkoláb: <i>Analysis of Include Dependencies in C++ Source Code</i>	4
Zsolt Bagóczki, Balázs Bánhelyi: <i>A parallelized interval arithmetics based reliable computing method on GPU</i>	5
Dénes Bán: <i>The Connection of Antipatterns and Maintainability in Firefox</i>	6
Áron Baráth, Gábor Alex Ispánovics, Porkoláb Zoltán: <i>Checking binary compatibility for modern programming language</i>	8
János Baumgartner, Zoltán Süle, János Abonyi: <i>Cox regression based process optimization</i>	9
Biswajeeban Mishra: <i>Improving OpenStack services with source code optimizations</i>	10
Bence Bogdándy, Zsolt Tóth: <i>Comparison of WiFi RSSI Filtering Methods</i>	11
Tibor Brunner, Norbert Pataki, Zoltán Porkoláb: <i>Finding semantical differences between C++ standard versions</i>	13
Ádám Budai, Kristóf Csorba: <i>RJMCMC Optimization of marble thin section image segmentations</i>	14
Márton Búr, Dániel Varró: <i>Towards hierarchical and distributed run-time monitors from high-level query languages</i>	16
Cristian Babau, Marius Marcu, Mircea Tihu, Daniel Telbis, Vladimir Cretu: <i>Real Time Road Traffic Characterization Using Mobile Crowd Sourcing</i>	17
Gábor Fábián: <i>Physical simulation based technique for determining continuous maps between meshes</i>	19
Tamás Fekete, Gergely Mezei: <i>Architectural challenges in creating a high-performant model-transformation engine</i>	20
Katalin Friedl, László Kabódi: <i>Storing the quantum Fourier operator in the QuIDD data structure</i>	22
Péter Gál, Csaba Bátori: <i>Bringing a Non-object Oriented Language Closer to the Object Oriented World: A Case Study for A+.NET</i>	23
Ábel Garai, István Péntek: <i>Common Open Telemedicine Hub and Interface Standard Recommendation</i>	24
László Gazdi, Dorottya Bodolai, Dr. Bertalan Forstner Ph.D., Luca Szegletes: <i>Supervising Adaptive Educational Games</i>	26
Péter Gyimesi: <i>An Automatic Way to Calculate Software Process Metrics of GitHub Projects With the Aid of Graph Databases</i>	27
Lajos Györffy, András London, Géza Makay: <i>The structure of pairing strategies for k-in-a-row type games</i>	29
László Hajdu, Miklós Krész, László Tóth: <i>Maximization Problems for the Independent Cascade Model</i>	30
Ferenc Horváth: <i>Eliminating code coverage differences from large-scale programs</i>	31

Laura Horváth, Krisztián Pomázi, Bertalan Forstner, Luca Szegletes: <i>Multiple Intelligence Learning</i>	32
István Kádár: <i>The Optimization of a Symbolic Execution Engine for Detecting Runtime Errors</i>	34
György Kalmár, László G. Nyúl: <i>Image processing based automatic pupillometry on infrared videos</i>	36
Melinda Katona, László G. Nyúl: <i>Quantitative Assessment of Retinal Layer Distortions and Sub-retinal Fluid in SD-OCT Images</i>	37
Krisztian Koós, Begüm Peksel, Lóránd Kelemen: <i>Height Measurement of Cells Using DIC Microscopy</i>	38
Levente Lipták, László Kundra, Dr. Kristóf Csorba: <i>Robust multiple hypothesis based enumeration in complex traffic scenarios</i>	39
Gábor Márton, Zoltán Porkoláb: <i>Unit Testing and Friends in C++</i>	40
Abigél Mester, Balázs Bánhelyi: <i>Revision of local search methods in the GLOBAL optimization algorithm</i>	41
Csaba Molnár, József Molnár: <i>Multi-layer phase field model for selective object extraction</i>	42
Balázs Nagy: <i>Skyline extraction by artificial neural networks for orientation in mountainous environment</i>	43
Teréz Nemes, Ákos Dávid, Zoltán Süle: <i>Decision-support system to maximize the robustness of computer network topologies</i>	44
Dávid Papp, Gábor Szűcs: <i>Active Learning using Distribution Analysis for Image Classification</i>	45
Zsolt Parragi, Zoltan Porkoláb: <i>Preserving Type Information in Memory Usage Measurement of C++ Programs</i>	47
Patrik J. Braun, Péter Ekler: <i>Improving QoS in web-based distributed streaming services with applied network coding</i>	48
Aleksandar Pejčić: <i>Detecting Similarities in Process Data generated with Computer-based Assessment Systems</i>	50
Tamás Pflanzner, Attila Kertész: <i>Taxonomy and Survey of IoT Cloud Use Cases</i>	52
Rebaz Ahmed Hama Karim: <i>Long-acting insulin management for blood glucose prediction model</i>	53
Selmeci Attila, Orosz Tamás Gábor: <i>Trends and paradigm change in ERP datacenter management</i>	55
Oszkár Semeráth, Dániel Varró: <i>Validation of Well-formedness Constraints on Uncertain Models</i>	57
Ferenc Attila Somogyi: <i>On the Efficiency of Text-based Comparison and Merging of Software Models</i>	58
Gábor Szőke: <i>Finding Matching Source Code Elements in an AST Using Position Information</i>	59
Judit Tamás, Zsolt Tóth: <i>Classification based Symbolic Indoor Positioning over the Miskolc IIS Dataset</i>	60
Máté Tömösközi: <i>Regression Model Building and Efficiency Prediction of Header Compression Implementations for VoIP</i>	62
Zsolt Vassy, István Kósa PhD, MD, István Vassányi PhD: <i>Stable Angina Clinical Pathway Correlation Clustering</i>	63
Dániel Zombori, Balázs Bánhelyi: <i>Parallel implementation of a modular, population based global optimizer package</i>	65
List of Authors	66

Program

Monday, June 27

08:00 – 08:50	Registration and Opening
08:50 – 10:00	Talks – Testing (3x20 minutes)
10:00 – 11:00	Talks – Mining (2x20 minutes)
11:00 – 12:40	Talks – Machine Learning (5x20 minutes)
12:40 – 14:00	Lunch
14:00 – 15:00	Plenary talk
15:00 – 16:40	Talks – Medical Solutions (4x20 minutes)
16:40 – 18:00	Talks – Program Analysis (4x20 minutes)
19:00	Reception at the Rector’s Building

Tuesday, June 28

08:30 – 09:40	Talks – Programming Languages (3x20 minutes)
09:40 – 11:00	Talks – Image processing I. (3x20 minutes)
11:00 – 12:00	Plenary talk
12:00 – 12:40	Talks – Model Matching (2x20 minutes)
12:40 – 14:00	Lunch
14:00 – 15:40	Talks – Image Processing II. (4x20 minutes)
15:40 – 17:00	Talks – Algorithm (4x20 minutes)
17:15	Social program
19:00	Gala Dinner at the Rector’s Building

Wednesday, June 29

08:30 – 09:40	Talks – Cloud/IoT (3x20 minutes)
09:40 – 11:00	Talks – Networking (3x20 minutes)
11:00 – 12:00	Plenary talk
12:00 – 12:40	Talks – Models (2x20 minutes)
12:40 – 14:00	Lunch
14:00 – 15:20	Talks – Optimization (4x20 minutes)
15:20	Closing

Detailed program

Monday, June 27

08:00	Registration
08:30	Opening
Session 1	Testing
08:50	Ferenc Horváth: <i>Eliminating code coverage differences from large-scale programs</i>
09:10	István Kádár: <i>The Optimization of a Symbolic Execution Engine for Detecting Runtime Errors</i>
09:30	Gábor Márton, Zoltán Porkoláb: <i>Unit Testing and Friends in C++</i>
09:50	Break
Session 2	Mining
10:00	Cristian Babau, Marius Marcu, Mircea Tihu, Daniel Telbis and Vladimir Cretu: <i>Real Time Road Traffic Characterization Using Mobile Crowdsourcing</i>
10:20	János Baumgartner, Zoltán Süle, János Abonyi: <i>Cox regression based process optimization</i>
10:40	Break
Session 3	Machine Learning
11:00	Laura Horváth, Krisztián Dániel Pomázi, Luca Szegletes, Bertalan Forstner: <i>Multiple Intelligence Learning</i>
11:20	Judit Tamás, Zsolt Tóth: <i>Classification based Symbolic Indoor Positioning over the Miskolc IIS Dataset</i>
11:40	Dénes Bán: <i>The Connection of Antipatterns and Maintainability in Firefox</i>
12:00	Aleksandar Pejic: <i>Detecting Similarities in Process Data generated with Computer-based Assessment Systems</i>
12:20	László Gazdi, Dorottya Bodolai, Bertalan Forstner, Luca Szegletes: <i>Supervising Adaptive Educational Games</i>
12:40	Lunch
14:00	Plenary Talk Dániel Marx: <i>The Optimality Program in Parameterized Algorithms</i>
14:50	Break
Session 4	Medical Solutions
15:00	Ábel Garai, István Péntek: <i>Common Open Telemedicine Hub and Interface Standard Recommendation</i>
15:20	Zsolt Vassy, István Vassányi, István Kósa: <i>Stable angina clinical pathway correlation clustering</i>
15:40	Melinda Katona, László G. Nyúl: <i>Quantitative Assessment of Retinal Layer Distortions and Subretinal Fluid in SD-OCT Images</i>
16:00	Rebaz Ahmed Hama Karim, István Vassányi, István Kósa: <i>Long-acting insulin management for blood glucose prediction model</i>
16:20	Break

Session 5	Program Analysis
16:40	Zsolt Parragi, Zoltán Porkoláb: <i>Typegrind: Type Preserving Heap Profiler for C++</i>
17:00	Péter Gyimesi: <i>An Automatic Way to Calculate Software Process Metrics of GitHub Projects With the Aid of Graph Databases</i>
17:20	Bence Babati, Norbert Pataki, Zoltán Porkoláb: <i>Analysis of Include Dependencies in C++ Source Code</i>
17:40	Biswajeeban Mishra: <i>Improving OpenStack Services with Source Code Optimizations</i>
<hr/>	
18:00	Free program
<hr/>	
19:00	Reception at the Rector's Building

Tuesday, June 28

Session 6	Programming Languages
08:30	Tibor Brunner, Norbert Pataki, Zoltán Porkoláb: <i>Finding semantical differences between C++ standard versions</i>
09:50	Áron Baráth, Zoltán Porkoláb: <i>Checking binary compatibility for modern programming language</i>
09:10	Péter Gál, Csaba Bátori: <i>Bringing a Non-object Oriented Language Closer to the Object Oriented World: A Case Study for A+.NET</i>
09:30	Break

Session 7	Image Processing I.
09:40	Balázs Nagy: <i>Skyline extraction by artificial neural networks for orientation in mountainous environment</i>
10:00	Dávid Papp, Gábor Szűcs: <i>Active Learning using Distribution Analysis for Image Classification</i>
10:20	Ádám Budai, Kristóf Csorba: <i>RJMCMC Optimization of marble thin section image segmentations</i>
10:40	Break

11:00	Plenary Talk László Pyber: <i>How to avoid the Classification Theorem of Finite Simple Groups in Asymptotic Group Theory</i>
11:50	Break

Session 8	Model Matching
12:00	Gábor Szőke: <i>Finding Matching Source Code Elements in an AST Using Position Information</i>
12:20	Ferenc Attila Somogyi: <i>On the Efficiency of Text-based Comparison and Merging of Software Models</i>

12:40	Lunch
-------	--------------

Session 9	Image Processing II.
14:00	György Kalmár, László G. Nyúl: <i>Image processing based automatic pupillometry on infrared videos</i>
14:20	Krisztián Koos, Begüm Peksel, Lóránd Kelemen: <i>Height Measurement of Cells Using DIC Microscopy</i>
14:40	Levente Lipták, László Kundra, Kristóf Csorba: <i>Robust multiple hypothesis based enumeration in complex traffic scenarios</i>
15:00	Csaba Molnár, József Molnár: <i>Multi-layer phase field model for selective object extraction</i>
15:20	Break

Session 10	Algorithm
15:40	Katalin Friedl, László Kabódi: <i>Storing the quantum Fourier operator in the QuIDD data structure</i>
16:00	Bence Bogdándy, Zsolt Tóth: <i>Comparison of WiFi RSSI Filtering Methods</i>
16:20	Lajos Györfy, András London, Géza Makay: <i>The structure of pairing strategies for \$-in-a-row type games</i>
16:40	Gábor Fábián: <i>Physical simulation based technique for determining continuous maps between meshes</i>

17:00	Free program
-------	--------------

17:15	Social Program Visit the Informatorium at Szent-Györgyi Albert Agóra
-------	--

19:00	Gala Dinner at the Rector's Building
-------	---

Wednesday, June 29

Session 11	Cloud/IoT
08:30	Tamás Pflanzner, Attila Kertész: <i>Taxonomy and Survey of IoT Cloud Use Cases</i>
08:50	Márton Búr, Dániel Varró: <i>Towards hierarchical and distributed run-time monitors from high-level query languages</i>
09:10	Attila Selmecsi, Tamás Orosz: <i>Trends and paradigm change in ERP datacenter management</i>
09:30	Break

Session 12	Networking
09:40	Teréz Nemes, Ákos Dávid, Zoltán Süle: <i>Proposing a decision-support system to maximize the robustness of computer network topologies</i>
10:00	Máté Tömösközi: <i>Regression Model Building and Efficiency Prediction of Header Compression Implementations for VoIP</i>
10:20	Patrik János Braun, Péter Ekler: <i>Improving QoS in web-based distributed streaming services with applied network coding</i>
10:40	Break

11:00	Plenary Talk Dániel Varró: <i>Models and Queries for Smart and Safe Cyber-physical Systems</i>
11:50	Break

Session 13	Models
12:00	Tamás Fekete, Gergely Mezei: <i>Architectural challenges in creating a high-performant model-transformation engine</i>
12:20	Oszkár Semeráth, Dániel Varró: <i>Validation of Well-formedness constraints on Uncertain Model</i>

12:40	Lunch
-------	--------------

Session 14	Optimization
14:00	Abigél Mester, Balázs Bánhelyi: <i>Revision of local search methods in the GLOBAL optimization algorithm</i>
14:20	Dániel Zombori, Balázs Bánhelyi: <i>Parallel implementation of a modular, population based global optimizer package</i>
14:40	Zsolt Bagóczki, Balázs Bánhelyi: <i>A parallelized interval arithmetics based reliable computing method on GPU</i>
15:00	László Hajdu, Miklós Krész, László Tóth: <i>Maximization Problems for the Independent Cascade Model</i>

15:20	Closing
-------	----------------

PLENARY TALKS

The Optimality Program in Parameterized Algorithms

Dániel Marx

Institute for Computer Science and Control, Hungarian Academy of Sciences (MTA SZTAKI), Hungary

Parameterized complexity analyzes the computational complexity of NP-hard combinatorial problems in finer detail than classical complexity: instead of expressing the running time as a univariate function of the size n of the input, one or more relevant parameters are defined and the running time is analyzed as a function depending on both the input size and these parameters. The goal is to obtain algorithms whose running time depends polynomially on the input size, but may have arbitrary (possibly exponential) dependence on the parameters. Moreover, we would like the dependence on the parameters to be as slowly growing as possible, to make it more likely that the algorithm is efficient in practice for small values of the parameters. In recent years, advances in parameterized algorithms and complexity have given us a tight understanding of how the parameter has to influence the running time for various problems. The talk will survey results of this form, showing that seemingly similar NP-hard problems can behave in very different ways if they are analyzed in the parameterized setting.

PLENARY TALKS

How to avoid the Classification Theorem of Finite Simple Groups in Asymptotic Group Theory

László Pyber

Hungarian Academy of Sciences, Rényi Institute (MTA Rényi), Hungary

The Classification of Finite Simple Groups (CFSG) is a monumental achievement and a seemingly indispensable tool in modern finite group theory. By now there are a few results which can be used to bypass this tool in a number of cases, most notably a theorem of Larsen and Pink which describes the structure of finite linear groups of bounded dimension over finite fields. In a few cases more ad hoc arguments can be used to delete the use of CFSG from the proofs of significant results. The talk will discuss a very recent example due to the speaker: how to obtain a CFSG-free version of Babai's quasipolynomial Graph Isomorphism algorithm by proving a weird lemma about permutation groups.

PLENARY TALKS

Models and Queries for Smart and Safe Cyber-physical Systems

Dániel Várró

Budapest University of Technology and Economics (BME), Hungary

A smart and safe cyber-physical system (CPS) autonomously perceives its operational context and adapts to changes over an open, heterogeneous and distributed platform with a massive number of nodes, dynamically acquires available resources and aggregates services to make real-time decisions, and resiliently provides critical services in a trustworthy way. In this talk, I present some challenges of CPS, and overview recent synthesis, exploration and validation techniques based on graph models and queries to assist the model-driven engineering of smart and safe CPS.

Analysis of Include Dependencies in C++ Source Code

Bence Babati, Norbert Pataki, Zoltán Porkoláb

The C++ Standard Template Library is the flagship example for libraries based on the generic programming paradigm. The usage of this library is intended to minimize classical C/C++ errors, but does not warrant bug-free programs [1]. Furthermore, many new kinds of errors may arise from the inaccurate use of the generic programming paradigm, like dereferencing invalid iterators or misunderstanding remove-like algorithms [4].

Unfortunately, the C++ Standard does not define which standard header includes another standard headers [4]. It is easy to write code that works perfectly on an implementation but fails to compile with another implementation of STL. These unportable codes should be result in compilation error with every STL implementation [3]. However, in this case the compiler does not warn us that this code is erroneous.

In this paper we present our tool that is based on the Clang [2]. This tool is able to detect the missing include directives that are patched by the STL implementation's internal structure. It also reports the unnecessary include directives to avoid extra compilation time. The background of our tool is discovered and we briefly present the underlying data structures and algorithms. We analyse how these problems occur in open source libraries and programs. Which environment proves oneself to be lazy or strict? How the developers take advantage of this portability issue?

References

- [1] Horváth, G., Pataki, N.: *Clang matchers for verified usage of the C++ Standard Template Library*, Annales Mathematicae et Informaticae, Vol. 44 (2015), pp. 99–109.
- [2] Lattner, C.: *LLVM and Clang: Next Generation Compiler Technology*, The BSD Conference, 2008.
- [3] Meyers, S.: *Effective STL*, Addison-Wesley, 2001.
- [4] Pataki, N.: *C++ Standard Template Library by Safe Functors*, in Proc. of 8th Joint Conference on Mathematics and Computer Science, MaCS 2010, Selected Papers, pp. 363–374.
- [5] B. Stroustrup: *The C++ Programming Language*, Addison-Wesley, special edition, 2000.

A parallelized interval arithmetics based reliable computing method on GPU

Zsolt Bagóczki, Balázs Bánhelyi

Videocards has outgrew in the past to be only a simple tool for graphic display. With their high speed video memories, lots of math units and parallelism, they provide a very powerful platform for general purpose computing tasks, in which we have to deal with large datasets, which are highly parallelizable, have high arithmetic intensity, etc. Our selected platform for testing is the CUDA (Compute Unified Device Architecture), that grants us direct reach to the virtual instruction set of the video card, and we are able to run our computations on dedicated computing kernels. The CUDA development kit comes with a useful toolbox with a wide range of GPU based function libraries.

In this parallel environment we implemented a reliable method. Hence the finite precision of computers, for achieving a reliable method, we needed to use interval arithmetics, so instead of giving a rounded, inaccurate result, we are able to give an interval, that is certainly including the exact answer. Our method is based on the branch-and-bound algorithm, with the purpose to decide whether or not any given property applies for a two-dimensional interval. This algorithm will give us the opportunity to use node level parallelization. Node level parallelization means, that our nodes are evaluated simultaneously on multiple threads we start on the GPU. This is called low-level, or type 1 parallelization, since we do not modify the searching trajectories, neither do we modify the dimensions of the branch-and-bound tree [1].

For testing, we choose the circle covering problem. It is the dual of circle packing, where our goal is to find the densest packing of a given number of congruent circles with disjoint interiors in a unit square. In circle covering, we aim for finding the full covering of the unit square with congruent circles of minimal radii. Overlapping interiors are allowed. For achieving easily scalable test cases, we will decide the covering with precalculated circles, where the diameter of the circles is the diagonal of the unit square, divided by a given number, n , and the number of the circles is the square of n . We scaled the problem up to three dimensions, and ran tests with sphere covering problems, too, and we discuss the possibility of scaling the problem up to any dimensions easily.

References

- [1] Teodor Gabriel Crainic, Bertrand Le Cun, and Catherine Roucairol: Chapter 1. Parallel Branch-and-Bound Algorithms, Parallel Combinatorial Optimization. Published Online (2006)

The Connection of Antipatterns and Maintainability in Firefox

Dénes Bán

The theory that antipatterns have a negative effect on source code maintainability is widely accepted but has relatively little objective research to support it. We aim to extend this research by analyzing the connection of antipatterns and maintainability in an empirical study of Firefox, an open source browser application developed in C++.

First, we selected 45 evenly distributed sample revisions from the `master` and `electrolysis` branches of Firefox between 2009 and 2010 – approximately one revision every two weeks. These provided the basis for both antipattern detection and maintainability assessment. We extracted the occurrences of 3 different antipattern types – Feature Envy, Long Function and Shotgun Surgery – defined by Brown et al. [1], using the same tool we previously published [2]. The antipatterns from the literature were interpreted in this case as violations of specific metric thresholds and/or structural constraints. E.g., for a method to be considered an instance of Feature Envy, it had to have more attribute accesses than a configurable limit while simultaneously the ratio of these accesses referring to foreign – i.e., non-local – attributes had to be higher than another limit. This would lead us to believe that the method itself is interested in attributes, but mainly not the ones belonging to its parent class, hence we call it “envious” of the other classes.

After extracting all the antipatterns, we summed the number of matches by type and divided them by the total number of logical lines of the subject system for each revision to create new, system-level antipattern density predictor metrics. This division makes sure that a system can be considered more maintainable than another even when it has more antipattern instances given that the ratio of these instances to the size of the system is less.

Next, we computed corresponding maintainability values using a C++ specific version of the probabilistic quality model published by Bakota et al. [3], which calculates increasingly more abstract source code characteristics by performing a weighted aggregation of lower level metrics according to the ISO/IEC 25010 standard [4]. Its final result is a number between 0 and 1 indicating the maintainability of the source code.

After this, we checked for correlations between the maintainability values and each of the different antipattern densities to unveil a connection between the underlying concepts. We found that Feature Envy, Long Function and Shotgun Surgery had coefficients of -0.95, -0.94 and -0.36 for Pearson’s, and -0.83, -0.85 and -0.57 for Spearman’s correlation, respectively. These figures show strong, inverse relationships, thereby supporting our initial assumption that the more antipatterns the source code contains, the harder it is to maintain.

Finally, we combined these data into a table applicable for machine learning experiments, which we conducted using Weka [5] and a number of its built-in algorithms. All regression types we tried for predicting source code maintainability from antipattern information – namely, Linear Regression, Multilayer Perceptron, REPTree, M5P and SMO Regression – reached correlation coefficients over .93 while using negative weights for the antipattern predictors in their built models. This not only means we can give precise estimations for the maintainability of the source code using its antipattern densities alone but also emphasizes their reversed roles.

In conclusion, we believe that this empirical study is another step towards objectively proving that antipatterns have an adverse effect on software maintainability.

References

- [1] Brown, W.J., Malveau, R.C., McCormick III, I.H.W., Mowbray, T.J.: *AntiPatterns: Refactoring Software, Architectures, and Projects in Crisis*. John Wiley & Sons, Inc., 1998.

- [2] Dénes Bán, Rudolf Ferenc: Recognizing Antipatterns and Analyzing their Effects on Software Maintainability, *In Proceedings of the 14th International Conference on Computational Science and Its Applications*, 2014.
- [3] Tibor Bakota, Péter Hegedűs, Péter Körtvélyesi, Rudolf Ferenc, and Tibor Gyimóthy: A Probabilistic Software Quality Model. *In Proceedings of the 27th IEEE International Conference on Software Maintenance*, 2011.
- [4] ISO/IEC: ISO/IEC 25000:2005. Software Engineering – Software product Quality Requirements and Evaluation (SQuaRE) – Guide to SQuaRE. 2005.
- [5] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten: The WEKA Data Mining Software: An Update. *SIGKDD Explorations*, 2009.

Checking binary compatibility for modern programming language

Áron Baráth, Gábor Alex Ispánovics, Porkoláb Zoltán

Modern programming languages prefer to support rapid development improved with flexible and expressive features. It seems languages take more effort in development rather than long-time support of the code. This trend can be seen in the fact that many languages handles binary compatibility poorly. In many cases only a very little change can cause serious runtime problems, like miscalculations and crashes.

In C [3] and C++ [4] (and in many other languages) binary compatibility starts when linking multiple objects into an executable or dynamic library. Using any build system, we can give situations when the built system will not recognize the dependencies correctly, and the compiler outputs a wrong binary – it is more likely when system-wide headers are also involved. Furthermore, the same issue can arise when linking against static libraries. These problems can be avoided with a local database of detailed information about the types and functions. The problem is getting more uncontrollable when a client uses dynamic libraries. While in C only the name of the functions and objects (also known as global variables) are used in static- and dynamic linking, in C++ *mangled names* are used to link functions. The mangled names will provide a little more validation when loading a library, but it is not nearly sufficient.

In the other hand, we have to calculate with the expected incompatibility: obviously, an interface modification will break the user programs, while modifications are not related to the interface itself are the most problematic. Many languages cannot handle correctly the second case due to optimization reasons or due to information loss.

Note that, the problem of the binary compatibility is not limited to languages like C and C++, but it is a real issue in managed languages as well – for example in Java language. Researches aimed to identify the possible weakpoints [1], others try to provide a solution for that [2].

In this paper we present an experimental method to detect binary incompatibilities in C++ program. Moreover we introduce our experimental programming language, Welltype [5], which is aimed to present a solution to detect unwanted binary incompatibilities at link-time. Our approach will not load incompatible programs into the same runtime context.

References

- [1] Dietrich, J., Jezek, K., Brada, P.: *Broken promises: An empirical study into evolution problems in java programs caused by library upgrades*. Software Maintenance, Reengineering and Reverse Engineering (CSMR-WCRE), 2014 Software Evolution Week-IEEE Conference on. IEEE, (2014)
- [2] Savga, I., Rudolf M., Goetz, S.: *Comeback!: a refactoring-based tool for binary-compatible framework upgrade*. Companion of the 30th international conference on Software engineering. ACM, (2008)
- [3] Kernighan, B. W., and Ritchie, D. M.: *The C programming language*. Vol. 2. Englewood Cliffs: prentice-Hall (1988)
- [4] Stroustrup, B. *The C++ Programming Language, 4th Edition*. Addison-Wesley (2013)
- [5] Welltype web page. <http://baratharon.web.elte.hu/welltype/>

Cox regression based process optimization

János Baumgartner, Zoltán Süle, János Abonyi

Data mining is an efficient tool to reveal information and to discover correlations derived out of this data. Process mining and event analysis become more and more popular research area for investigating a sequence of events. Examining the available set of information leads straight to the opportunity of either iterative process development or optimization. According to the general interpretation of survival analysis [1][2] a process can be investigated by focusing on a special event of interest, thus an estimation can be given for the expected duration of surviving time. The Cox's proportional hazard model is capable to divide the entire investigation period into spells (i.e. sub periods) [3]. Thus, the possibility for investigation of different sub periods can be ensured, so the entire process is influenced by each sub process in every well-defined time slot, hence the overall risk and the shape of the survival or hazard function also differs from time to time. We introduce a novel methodology by which time and also cost can be saved by determining the optimal sequence of sub processes in the considered process. Contrary to the classical survival analysis the core idea is to examine the data of a test process consisting of sub process steps, and based on the gained information the sequence of these sub elements can be redesigned. Additionally, when parameters also have to be taken into account, the result of the investigation is affected accordingly. Therefore, parametrical survival patterns can be fitted to the problem, so the risks for each time period can be determined. Using Cox regression we can highlight those process steps including those relevant parameters which increase significantly the risk. It is also important to emphasize that the fault of a process step not exclusively means the fault of the entire process, it assumes a rising risk of the overall fault. These individual sub hazard functions assigned to the different time periods build up a complex characteristic survival function of the regarded process. As a further development we inspect the possibilities of extending the basements as a problem class in connection with neural networks. The results are illustrated through a realistic example taken from manufacturing and analysis of education data.

Acknowledgements

This publication has been supported by the Hungarian Government through the project VKSZ_14-1-2015-0190 - Development of model based decision support system for cost and energy management of electronic assembly processes. The research of Janos Abonyi has been supported by the National Research, Development and Innovation Office - NKFIH, through the project OTKA - 116674 (Process mining and deep learning in the natural sciences and process development).

References

- [1] Brenner, H., Castro, F., Eberle, A., Emrich, K., Holleczech, B., Katalinic, A., Jansen, L. (2016), *Death certificate only proportions should be age adjusted in studies comparing cancer survival across populations and over time*, European Journal of Cancer, 52, 102-108.
- [2] Ross, A., Matuszyk, A., Thomas, L. (2010), *Application of survival analysis to cash flow modelling for mortgage products*, OR Insight, 23, 1-14
- [3] Willett, J., Singer, J., (1993), *Investigating Onset, Cessation, Relapse, and Recovery: Why You Should, and How You Can, Use Discrete-Time Survival Analysis to Examine Event Occurrence*, Journal of Consulting and Clinical Psychology, 952-965

Improving OpenStack services with source code optimizations

Biswajeeban Mishra

Cloud Computing is a broad research area that uses many aspects of software and hardware solutions, including computing and storage resources, application runtimes or complex application functionalities. Clouds enable the outsourcing of IT infrastructure management for users, allowing them to concentrate on their core competencies to get better performance [1].

OpenStack is an open source cloud computing framework having a built-in modular architecture, based on the IaaS (Infrastructure as a Service)[2] model. It was founded in 2010, currently it is managed by the OpenStack Foundation, a non-profit corporate entity established in September 2012. Since then more than 500 companies have joined the project. It has rapidly grown into a global software community of developers and cloud computing technologists collaborating on an open-source cloud operating system for both public and private clouds. OpenStack clouds are powered by a series of interrelated projects to support virtualized infrastructure and application management forming a robust and complex distributed system [3]. All these services are controlled and managed using Openstack's Dashboard, a web based graphical user interface, which provides administrators and users to access, provision, and automate cloud-based resources.

The aim of this paper is to introduce source code analysis on OpenStack with a predefined tool-chain (including a coding rule checker, metrics calculator, and duplicated code detector), and to discuss how to derive concrete blueprints targeting improvements of the internal quality of the platform. This process requires contributions from both theoretical research and practical development, including software tool releases and prototypes, tracking and fixing known issues and defects in OpenStack software by understanding the complete development work flow and architecture.

References

- [1] Buyya, Rajkumar and Yeo, Chee Shin and Venugopal, Srikumar and Broberg, James and Brandic, Ivona, *Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility*, Future Generation computer systems, vol. 25, no. 6, pp. 599-616, 2009
- [2] Sefraoui, Omar and Aissaoui, Mohammed and Eleuldj, Mohsine, *OpenStack: toward an open-source solution for cloud computing*, International Journal of Computer Applications, vol. 55, no. 3, 2012
- [3] Teixeira, Jose, *Understanding coopetition in the open-source arena: The cases of webkit and openstack*, Proceedings of The International Symposium on Open Collaboration, pp. 39, 2014

Comparison of WiFi RSSI Filtering Methods

Bence Bogdándy, Zsolt Tóth

Indoor Positioning Systems have gained attention in the last decade. WLAN is one of the most popular technology among fingerprinting based Indoor Positioning Systems. RADAR was the first fingerprinting based system that recorded the WIFI RSSI value at known location. The HORUS system [1, 7] is location determining system [3] based on the IEEE 802.11b WiFi connection. It showed that using time series instead of single values increases accuracy of indoor positioning algorithms. The system used a simple filter which was taking the average of the measurements.

The applied use of filters [5] as preprocessing algorithms on ongoing measurements to separate data from noise is an important research area. The usage of such filters [2] can provide a data set free of inconsistencies. The filtered data can be used to accurately describe a location. This paper focuses on the analysis of time series [8] of WiFi RSSI measurements. The analysis of the time series could lead to the development of an efficient client site filtering method for indoor positioning systems [4, 6]. The Wifi RSSI values were recorded in order to perform the efficiency analysis of the Filters. The recording took place in the University of Miskolc's Institution of Information Science, Department of Information Technology. The values were used to create a data set in order to test the filters. The paper focuses on the comparison of three different filtering methods. The first filter is HORUS' algorithm which takes the average of the last m values where m is the memory size. The second filter is also based on time windowing but it has two parameters threshold and a memory size. The algorithm calculates the total difference of the values in the memory. If the sum exceeds the threshold, then the current value is replaced by the average of the values in the memory. The third filter differs from the second one because it calculates the threshold dynamically. Thus, this filter requires less parameters and could be more adaptive. These filters were tested and compared over the pre-recorded dataset. R was used to implement the tested filtering methods that results are shown in Figure 1. These line charts show how the RSSI values (dB) changes over a time (s) period. Figure 1(a) shows the unfiltered dataset that range was up to -60 dB. The HORUS' filter could decrease the range to -15 dB that is shown in Figure 1(b). Moreover, the filtered RSSI values become real number due to the usage of average function. Figure 1(c) shows the results of the filter that uses a predefined threshold. The filtered values are not necessarily replaced with the average so the measured values can be contained in the yield data. So this method can be used to eliminate the random observation errors caused by temporal events. Figure 1(d) presents the results of the time windowing with dynamic threshold. However this method have larger range than the previous filters, but it depends on only one parameter which eases its usage. The presented comparison of filtering methods allows us to create efficient client side filtering to improve stability of the measured RSSI values. The filters create a more stable series where the range of values is reduced. Usage of client side filtering can improve the accuracy of the Indoor Positioning System. These filters will be implemented on Android platform in order to test them in real life scenarios and to test their effect on the ILONA system.

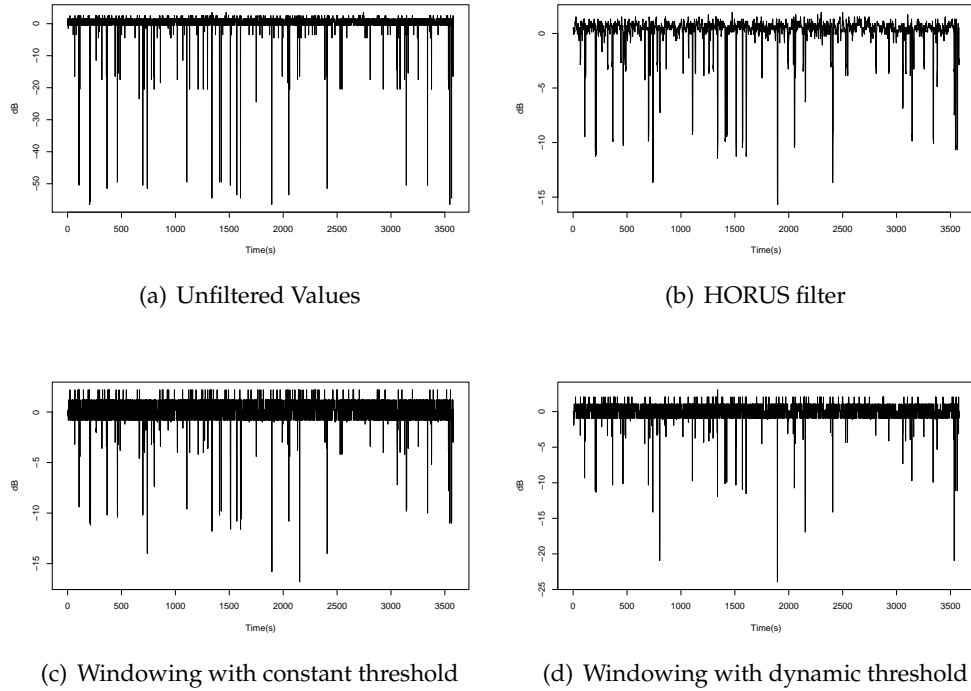


Figure 1: Results of the tested filtering methods

References

- [1] The horus wlan location determination system. https://www.usenix.org/legacy/publications/library/proceedings/mobisys05/tech/full_papers/youssef/youssef_html/index.html.
- [2] Properties of indoor received signal strength for wlan location fingerprinting.
- [3] Artur Baniukevic, Christian S Jensen, and Hua Lu. Hybrid indoor positioning with wi-fi and bluetooth: Architecture and performance. In Mobile Data Management (MDM), 2013 IEEE 14th International Conference on, volume 1, pages 207-216. IEEE, 2013.
- [4] Yongguang Chen and Hisashi Kobayashi. Signal strength based indoor geolocation. In Communications, 2002. ICC 2002. IEEE International Conference on, volume 1, pages 436-439. IEEE, 2002.
- [5] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. Journal of basic Engineering, 82(1):35-45, 1960.
- [6] Richárd Németh Judit Tamás Zsolt Tóth, Péter Magnucz. Data model for hybrid in-door positioning systems. Production Systems and Information Engineering, 7(1):67-80, 12 2015. HU ISSN 1785-1270.
- [7] Moustafa Youssef and Ashok Agrawala. The horus wlan location determination system. In Proceedings of the 3rd international conference on Mobile systems, applications, and services, pages 205-218. ACM, 2005.
- [8] William Wu-Shyong Wei. Time series analysis. Addison-Wesley publ Reading, 1994.

Finding semantical differences between C++ standard versions

Tibor Brunner, Norbert Pataki, Zoltán Porkoláb

Programming languages are also participants of the modern software technology. Software systems are getting increasingly large-scale, which means that their code base also gets more and more complex. To compensate this, programming languages are providing high-level constructions to ease the handling of higher abstraction levels.

C++ is one of the most widespread programming languages among those applications which require high runtime performance. In the last few years C++ has introduced new features to enable creation of faster programs with less code. Among others C++11 handles multi-threading for parallel tasks, and move semantics for administration of temporary objects to save memory and CPU usage.

Programmers' expectation is that they can benefit from these new language features, and their previously written codes have the same behaviour when a new compiler version is used. Unfortunately this is not the case. We have found some examples when building the same program with a new compiler may have different results. This is caused by the semantic change of some constructions between two versions of the language standard. This means that it can be a challenging task to maintain the huge amount of legacy code, developed over the years, and to ensure that their behaviour is unchanged after an update of the build environment.

In this article we draw the attention on some of the semantic changes between the different versions of the standard. We present a static analysis method for finding these differences, and provide a tool for listing them. For this purpose we use LLVM/Clang compiler infrastructure.

RJMCMC Optimization of marble thin section image segmentations

Ádám Budai, Kristóf Csorba

The project GrainAutLine [1] addresses the issue of easing the analysis procedure of marble thin section images which usually involves time consuming manual work. The traditional and the actually widespread method to extract valuable information about the images starts by drawing the borders of the grains by hand. These are then used to produce grain size histograms and further statistics about the grains and their neighborhood. Experts use these results to draw conclusions about the provenance of the sample, the quality, and a lot of other properties. The project is about creating a software especially for geologists which enables them to access the required information in an automatic way (further endeavor) or at least semi-automatic way (closer purpose). The automation involves the following features: identifying the grains using image segmentation methods, and extracting statistic related information. The full automation is difficult due to the fact that marble grains are not artificial formations and they can show anomalies even an expert may be confused about. To eliminate uncertainty caused by the difficult cases the program lets the user make some interventions and correct mistakes. In this work a sub-project of GrainAutLine is shown which deals with only the image segmentation.

The proposed algorithm presented in this paper is for the segmentation of the grains in a special situation when a lot of twin crystals pose a challenge for the traditional algorithms like adaptive thresholding. A twin crystal is the result of more small crystals when they melt into a bigger one under special circumstances, as this leads to very characteristic patterns inside the grains which need to be distinguished from the true grain boundary lines. The basic idea is the over-segmentation of the image with another algorithm which is already available in the software. The over-segmented picture contains many little segments called blobs and the algorithm structures the blobs into bigger groups called superblobs. The concrete superblobs give a possible configuration of the system. The goal is to change this configuration until the superblobs match the real grains. The change occurs according to an energy function which depends on the configuration. The better the configuration, the smaller the total energy. [2] The capability of the algorithm revolves around the good choice of the energy function. Finally, the remaining task is the minimization of the energy function.

The energy function is the sum of four energy terms. Each term grabs a concrete characteristic of the shape of the grains: (1) the fact that grains tend to be convex, that (2) the straight parallel lines indicate twin crystals. (3) The actual technique relies on the user's intervention in some extent because the user is required to mark every grain in the input picture, and that (4) grains do not entirely surround other grains. The proposed algorithm uses the Reversible-jump Markov Chain Monte Carlo (RJMCMC) optimization method [3] in order to change the configuration of the superblobs until the energy function reaches its minimum point. The algorithm was tested on both artificial and real images and the profile of its performance in terms of processing time, accuracy and convergence speed. The results have proven that the concept of this algorithm can solve the segmentation of marble thin sections even in the presence of twin crystals.

Acknowledgements

This work has been supported by the Department of Automation and Applied Informatics at Budapest University of Technology and Economics.

References

- [1] Kristóf Csorba, Lilla Barancsik, Balázs Székely, Judit Zoldfoldi (2015). A Supervised Grain Boundary Extraction Tool Supported by Image Processing and Pattern Recognition, *ASMOSIA XI. International Conference* .
- [2] Yann Lecun et. al. (2006). A tutorial on energy based learning, *MIT Press*.
- [3] C. M. Bishop (2006). Pattern Recognition and Machine Learning, *Springer*.

Towards hierarchical and distributed run-time monitors from high-level query languages

Márton Búr, Dániel Varró

Context. A *cyber-physical system (CPS)* consists of computation, communication and control components tightly combined with physical processes of different nature, e.g., mechanical, electrical, and chemical [1]. Compared to traditional embedded systems, key characteristics of a CPS include a (1) *massive number of heterogeneous nodes* ranging from cheap, low-energy smart devices to mobile phones to high-end cloud-based servers, (2) *adaptability to conditions* that differ significantly from the ones they were designed for (new requirements, new services, new platforms, failures) in dynamic environments, while (3) *delivering critical services in a trustworthy way*. Such systems include autonomous and connected cars, smart healthcare devices, smart factories, smart homes or smart cities.

Problem statement. Due to their dynamic nature, the assurance of smart and trusted CPS typically relies on run-time verification, which aims to check if their execution at run-time meets its requirements [2]. For instance, the data provided by force torque sensors and tactile sensors of an automated robot arm can be evaluated to decide if the arm is in a dangerous situation. High-level property languages are increasingly used for specifying complex structural conditions of the system. These properties are either evaluated over run-time models (which are directly connected to the system itself) or they serve as an input for synthesizing run-time monitors for a heterogeneous platform. However, due to resource constraints of these devices (e.g. CPU, memory, energy) and the continuously evolving platform and services, the deployment of hierarchical monitors to such a target platform is a very challenging task.

Objectives. In this line of research, we aim to continuously evaluate properties captured in high-level query languages over run-time models by deploying them over a heterogeneous platform for run-time verification purposes. While efficient incremental and search plan-based query optimization techniques have been developed in model-driven engineering [3] as well as for graph databases [4], their application in an environment with strict resource constraints and soft real-time requirements is a major challenge. As a first step, we illustrate these challenges in the context of the MoDeS3 demonstrator [5] developed for the Eclipse IoT Challenge 2016.

References

- [1] Baheti, Radhakisan, and Helen Gill. "Cyber-physical systems." *The impact of control technology* 12 (2011): 161-166.
- [2] Medhat, Ramy, et al. "Runtime Monitoring of Cyber-Physical Systems Under Timing and Memory Constraints." *ACM Transactions on Embedded Computing Systems (TECS)* 14.4 (2015): 79.
- [3] Horváth, Ákos, Gergely Varró, and Dániel Varró. "Generic search plans for matching advanced graph patterns." *Electronic Communications of the EASST* 6 (2007).
- [4] Schmidt, Michael, Michael Meier, and Georg Lausen. "Foundations of SPARQL query optimization." *Proceedings of the 13th International Conference on Database Theory*. ACM, 2010.
- [5] MoDeS3 Project. <http://inf.mit.bme.hu/en/research/projects/modes3>

Real Time Road Traffic Characterization Using Mobile Crowd Sourcing

Cristian Babau, Marius Marcu, Mircea Tihu, Daniel Telbis, Vladimir Cretu

Traffic optimization is a subject that became vital for the world we live in. People need to get from a starting point to a destination point as fast and safe as possible. Traffic congestion play a very important role in the frustration of people resulting in time loss, reduced productivity and wasted resources. With our research we aim to address these issues by proposing a real time road traffic planning system based on mobile context and crowd sourcing efforts. The first step toward this goal is real-time traffic characterization using data collected from mobile sensors of the traffic participants: drivers, pedestrians, cyclists, passengers. We started developing a data collection and analysis system composed of a mobile application to collect user context data and a web application to view and analyze the data.

The architecture presented in Fig. 1 is composed of modules representing the main characteristics and functionality of the system. The data collection module can be further divided into two categories: the mobile sensors and the mobile local database. It collects environmental context information from sensors of the smart phones. It uses the accelerometer as the starting point of the system. Because of its low power consumption, the accelerometer could be always on. When it detects acceleration greater than a threshold it activates other sensors in the mobile probe depending on the momentary context. The location is gathered via Mobile Station Sensors, GPS and GSM. When the mobile device, for some reason, can't connect to the main server, then proposed system uses a database in the mobile device as a temporary cache database. Here the information about the environmental context is stored and transmitted to the server when the connection is available.



Figure 1: Overall architecture

The data transmission module uses the standard mobile connections available 3G/4G or Wi-Fi. When the connection is severed intentionally (user turns off data connection) or not (areas with bad data connection) then the environmental context data will be stored in the mobile database. When the connection is available, then the transmission of data resumes. Not all data is relevant or necessary or it is private. Initial data filtering could be done in the mobile phones to reduce the amount of data transmitted to the server. Then the server takes the data, analysis it, processes it and in this process filters it. Anonymization is a very important part of the whole process. The user's private information should not be published or transmitted in any way.

Another way of collecting the data is by using Social Media Crowd Sourcing. Social media like Facebook or Twitter are also used to collect traffic information. These techniques are not yet integrated into any ITS, because, like the radio station collection, the opinion regarding the status of the roads, congestion, incidents are subjective to the crowd. Popular route planning systems, like Google Maps, Bing Maps, Yahoo Map, generate driving directions using a static database of historical data [1]. They are built in the assumption of constancy and universality, the notion that an optimal route is independent of the weekday and the time of day [1]. In a number of countries Google Maps has a system called Google Traffic integrated into the Maps application that uses crowd sourcing as the main source of data collection [2]. Another route planning system is Waze which uses the subjective information given by the users regarding the status of the roads, congestion, incidents. Waze is using human users in the loop for large-scale sensing and computing, including human-powered sensing, human-centered computing,

transient networking, and crowd data processing and intelligence extraction [3]. Most of the web maps mentioned before give real time traffic information for a few road segments. However, the real time traffic conditions are just for the user information and it is not integrated into the driving route that the user queried. The suggested routes are still static, they do not change according to the weekday or the time of day or the congestion that could form, calculated by knowing the distance of two points in a graph and the legal speed a vehicle [4].

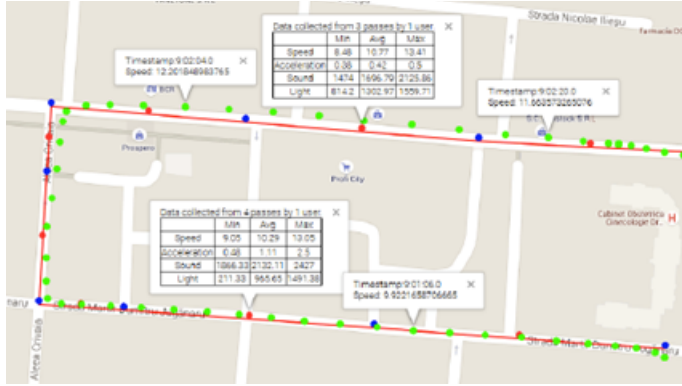


Figure 2: Tool overview

Each route recorded was parsed: for each segment provided by the API, a local query was used to determine if there are points which could be mapped to the API segment and then based on their timestamp the API segment end-points timestamps are estimated (Fig. 2). Then, based on the timestamps, the acceleration, speed, light intensity and sound intensity for the current API segment are estimated based on the sensor data collected. If a segment has no points that can be mapped inside of it, that the segment is not estimated for that particular route and may be estimated when analyzing a different route that may have locations that can be mapped inside the segment. We started the collection of context data in January 2016 in Timisoara, Romania. We now have 2.9 GB of data in the database. That includes over 17 million entries from accelerometer, over 58.000 entries from battery level, over 1600 entries from contexts, over 50.000 entries from foreground application, over 9 million entries from gyroscope, over 2.5 million entries from light level data, over 5.5 million entries from linear acceleration data, over 400.000 entries from GPS, over 9 million entries from magnetic field data, over 1.3 million entries from pressure data, over 220.000 entries from sound data, 321 processed routes, over 15.000 map segments.

References

- [1] J. Letchner, J. Krumm, E. Horvitz, "Trip Router with Individualized Preferences (TRIP): Incorporating Personalization into Route Planning", 2006.
- [2] F. Lardinois, "Google Maps Gets Smarter: Crowd sources Live Traffic Data", 2009.
- [3] B. Guo, Z. Wang, Z. Yu, Y. Wang, N.Y. Yen, R. Huang, X. Zhou, "Mobile Crowd Sensing and Computing: The Review of an Emerging Human-Powered Sensing Paradigm", ACM Computing Surveys, Vol. 48(1), 2015
- [4] Ji Yuan, Yi Zheng, Xi Xie, G. Sun, "T-Drive: Enhancing Driving Directions with Taxi Drivers' Intelligence", IEEE Transactions on Knowledge and Data Engineering, Vol. 25(1), 2013.

Physical simulation based technique for determining continuous maps between meshes

Gábor Fábián

The classification theorem of closed surfaces states, that any connected compact orientable surface is homeomorphic to the sphere or connected sum of n tori, for $n \geq 1$. If $X, Y \subset \mathbb{R}^3$ are topologically equivalent surfaces, then there exists an $H : X \rightarrow Y$ bijection between the two surface such H and H^{-1} are continuous functions. Let us suppose, that f and g are the parameterizations of X and Y are given, i.e. for some $I \subset \mathbb{R}^2$ two-dimensional interval $f : I \rightarrow X$ and $g : I \rightarrow Y$ bijections are known. The so-called homotopy between the functions f and g is defined as an $h : [0, 1] \times I \rightarrow \mathbb{R}^3$ continuous map, such $h(0, x) = f(x)$ and $h(1, x) = g(x)$. As f and g are parameterization of the topologically equivalent surfaces, then h can be interpreted as a continuous deformation between the surface X and Y . In our research we are looking for this type of transition function for given X and Y surfaces. In computer graphics a surface is mostly represented by a triangular mesh, i.e. a set of vertices and the vertex indices of triangular surface elements. Since the parameterization of the surface is unknown, it is not an easy to define a transition between X and Y . The fact, that a given surface is topologically equivalent to the sphere (or n -tori) can be checked quite easily, but commonly defining the homeomorphism between surfaces or the homotopy between parameterizations is a difficult task. On the other hand, if we knew the parameterization of the surfaces, defining a homeomorphism would be much easier. Therefore we invented a method, that can determine the parameterization of meshes that are homeomorphic to the sphere. The main idea came from the real world, we simulate the motion of the surface particles of an air balloon, because if the internal pressure is large enough, the balloon surface is similar to a sphere. Let us consider the vertices, edges and the mesh to be particles, springs and surface of a soft body, respectively. Then we can simulate the pressure acting in the direction of face normals. As the faces move outwards, the volume of the body increases, therefore the pressure decreases. The springs do not let the adjacent vertices to move far away from each other. Obviously, if each vertices displaced to a sphere, then the spring force and pressure force act in the opposite direction, therefore this is a state of equilibrium. If the simulation is successful, then we get a homeomorphism between a given surface and the sphere, therefore we can easily determine the parameterization and the transition function between surfaces. We will present our results in physical simulation, mesh parameterization and determining homeomorphisms and transitions between meshes.

References

- [1] K. Crane, U. Pinkall, P. Schröder. Robust Fairing via Conformal Curvature Flow, *ACM Transactions on Graphics (TOG) Volume 32 Issue 4*, 2013.
- [2] C. Ericson. Real-Time Collision Detection. *Elsevier*, 2005.
- [3] G. Fábián, L. Gergó. Planning data structure for space partitioning algorithms, *International Journal of Applied Mathematics and Informatics*, revised, 2016.
- [4] S. Ghali. Introduction to Geometric Computing, *Springer-Verlag London Limited*, 2008.
- [5] J. M. Lee. Introduction to Topological Manifolds, *Springer*, 2000.

Architectural challenges in creating a high-performant model-transformation engine

Tamás Fekete, Gergely Mezei

Modeling is a frequently used concept in computer science. One of its main application areas is the Model-driven Engineering (MDE). Due to the essential role of model processing in MDE, it is important to find and apply efficient model-transformation algorithms. Several techniques exist, one of the most popular among them is the graph rewriting-based model transformation. Graph rewriting is based on subgraph isomorphism, which is an NP complete problem. Since MDE tools must often handle huge models, performance is a key feature. However, graph rewriting tends to slow down sharply as the size of the input model increases. The usage of strongly parallelized algorithms is a way to increase the performance. Nowadays, there are several kinds of hardware components supporting efficient evaluation of parallelized algorithms. OpenCL framework is the most popular interface supporting heterogeneous architectures (e.g. CPU, GPU, APU, DSP) [1]. The comparison of the OpenCL framework and other vendor specific solutions is studied in many papers e.g. in [2]. According to the classification of graph transformation approaches [3], none of them can use the advantages of the OpenCL framework. Our overall goal is to fill in this gap and create a high-performant general model-transformation engine based on the OpenCL framework. The engine we are creating is referred to as the GPGPU-based Engine for Model Processing (GEMP). In [4] we have worked out and evaluated the first part (pattern matching) of model transformations. However, model transformation is not complete without the ability of rewriting the pattern. In this paper, we take this step and describe the desinging and implementation challenges and our solution. The performance is studied on experimental way using a case study.

OpenCL versions are backwards compatible. Since NVIDIA GPUs support OpenCL 1.2 only, we have decided to use this version in order to maximize the hardware independence. The implementation is applied in C++11 (STL and Boost libraries are used), we are using the test-driven development methodology (TDD). The evaluation of the unit tests provided immediate feedback about the state of the implementation continuously from the beginning, which was a great help in judging our progress. **As the main goal is to achieve a high-performant model-transformation engine, the following four main points are discussed in the current paper:** (i) The description of an efficient OpenCL-based pattern matching algorithm. (ii) As far as we use several devices, we need to synchronize their work in order to achieve the best performance (e.g.: CPU + GPU). The synchronization algorithm is elaborated. (iii) Scalability is studied because of the limited memory usage (e.g.: dividing the input/output data). (iv) The data structures and the relations between those are also highly important because of the high performance and low memory usage.

The novelties of the current paper are the followings: (i) The detailed architecture of a high-performant model-transformation engine supporting both pattern matching and rewriting. (ii) A collection of acceleration techniques which are applied and (iii) optimization models to find the best configuration for the model-transformation. The efficiency of our solution is proved through case studies. Results are introduced and evaluated in the paper.

In general, model-transformation consists of three main steps: (i) Finding topological matches in the graph fulfilling both the negative and the positive constraints of the pattern. (ii) Processing the attributes of the result of the first step and thus evaluating the constraints referring to attributes. (iii) Rewriting the pattern. The first two steps are executed on OpenCL devices, while the third is executed directly on the host.

If the input model must be divided into partitions then the whole process is repeated as soon as processing of the partition of the input data is finished. Vertices, which cannot be processed on the OpenCL device (because of the partitions) are queued and evaluated on the host as soon as the host has free capacity.

Both the positive and negative constraints are evaluated together in two different steps. The reason for dividing the matching algorithms into two steps is that copying all vertices of the host model with all of their attributes to the OpenCL device memory would be inefficient. By dividing this process into two strongly connected steps, we can reduce the data to be copied. In this way, we need to copy the attributes only for those vertices, which are part of a topological matching structure. In the first two steps in GEMP, there are two separated kernel codes. The third step of graph rewriting is the rewriting of the graph, namely replacing the pattern found in the input graph as specified by the rewriting rule. This step is managed on the host side, without the usage of any OpenCL device. This is necessary, since matches may overlap and threads are working in parallel. This means that the patterns found by the OpenCL devices may be invalid at this time (invalidated by other threads). Therefore, the host acts as a synchronization point and ensures the validity of the patterns before rewriting.

Our overall goal is to create a GPU-based model transformation engine referred to as GEMP. The main contribution of the paper is the rewriting part of the transformation. This part is elaborated in detail. Moreover, the paper presents the architecture of the solution and illustrates the mechanisms, analyzes the results by case studies. As the domain for the case studies, the movie database, IMDB [5] was chosen. Although our goal is not fully complete, the preliminary results are promising which means that GEMP can be effectively used as an accelerator engine in all MDE tools.

References

- [1] K. Group. [Online]. Available: <https://www.khronos.org/opencl/>. [Accessed 15 January 2016].
- [2] Veerasamy, Bala Dhandayuthapani; Nasira, G. M.. *Exploring the contrast on GPGPU computing through CUDA and OPENCL*, Journal on Software Engineering vol. 9, issue 1, p. 1-8, (2014).
- [3] E. Jakumeit, S. Buchwald, D. Wagelaar, L. Dan, A. Hegedus, M. Herrmannsdorfer, T. Hornf, E. Kalnina, C. Krause, K. Lano, M. Lepper, A. Rensink, L. Rose, S. Watzoldt and S. Mazanek, *A survey and comparison of transformation tools based on the transformation tool contest*, Special issue on Experimental Software Engineering in the Cloud(ESEiC), vol. 85, no. 1, pp. 41-99, 2014.
- [4] F. Tamás and M. Gergely, *Creating a GPGPU-accelerated framework for pattern matching using a case study*, in EUROCON 2015, International Conference on Computer as a Tool, Salamanca, Spain, 2015.
- [5] *IMDb database source, Alternative Interfaces*, [Online]. Available: www.imdb.com/interfaces. [Accessed 15 January 2016].

Storing the quantum Fourier operator in the QuIDD data structure

Katalin Friedl, László Kabódi

One of the most known results of quantum computing is Shor's algorithm [2]. With a quantum computer running Shor's algorithm one can factor a composite number in expected polynomial time. But there is no known working quantum computer with more than a few qubits, so the algorithm is only a theoretical breakthrough.

Quantum algorithms can be simulated using classical computers, but the time complexity of the simulation is exponential. There are some data structures which can accelerate this. Viamontes et. al. described the QuIDD data structure [3] that is an extension of algebraic decision diagram, and wrote a quantum circuit simulator employing it, the QuIDDPro [4]. Using the QuIDD software one can simulate various quantum algorithms, for example Grover's famous search algorithm [1] in linear time. But the quantum operators used by Grover's algorithm differ from the ones used by Shor's factorization algorithm. The main difficulty in simulating Shor's algorithm is the quantum Fourier operator.

In this paper we examine the matrix of the Fourier operator, and its QuIDD representation. In its usual representation the decision diagram of an n qubit Fourier operator consists of approximately 2^{2n} nodes. There is an internal structure of the matrix, which we try to extract. First, we examine changing the ordering of the underlying decision diagram. This way we can use the recursive nature of the Fourier operator to achieve a diagram approximately $\frac{2}{3}$ times the size of the standard ordering. Then we try to emulate this change using permutation matrices, so the existing softwares can be used with greater efficiency. There are two permutation matrices we examine. One emulates the reverse ordering of the column variables, the other separates the columns based on their parity. Using the first one, we emulate the reversed ordering exactly, but the QuIDD representation of the permutation matrix is exponentially large. The second one exposes an internal symmetry, which can produce a diagram approximately $\frac{3}{4}$ times the original, but we can store the permutation matrix in linear space. Finally we propose a new method of storing the Fourier operator, using multipliers stored on the edges of the decision diagram. This helps reducing the size of the operator, but we do not know how the operations can be efficiently transferred to this variant.

Acknowledgements

This research is partially supported by the National Research, Development and Innovation Office (NKFIH), by the OTKA-108947 grant.

References

- [1] Lov K. Grover, A fast quantum mechanical algorithm for database search, Proceedings of the Twenty-eighth Annual ACM Symposium on Theory of Computing, 1996. <http://arxiv.org/pdf/quant-ph/9605043>
- [2] Peter W. Shor, Algorithms for Quantum Computation: Discrete Logarithms and Factoring, Proceedings of the 35th Annual Symposium on Foundations of Computer Science, 1994.
- [3] George F. Viamontes, Igor L. Markov, John P. Hayes, Improving Gate-Level Simulation of Quantum Circuits, Quantum Information Processing 2, 2003.
- [4] George F. Viamontes, Igor L. Markov, John P. Hayes, QuIDDPro: High-Performance Quantum Circuit Simulation, <http://vlsicad.eecs.umich.edu/Quantum/qp/>

Bringing a Non-object Oriented Language Closer to the Object Oriented World: A Case Study for A+.NET

Péter Gál, Csaba Bátori

In our previous work, we have introduced the A+.NET project [1], a clean room implementation of the A+ runtime for the .NET environment. In that implementation, we have kept the original A+ behaviour[2] and placed it on top of another runtime environment. This allowed us to execute A+ scripts which were written for the original engine without modification.

With the .NET based runtime there was already a way of accessing .NET objects from A+. However for each of the classes, methods, and variables which we wanted to access from A+, we needed to write lots of wrapper methods in C#. Writing these methods was usually a tedious task as they mostly followed the same structure.

In this work, we present an extension that allows us to provide a convenient way to handle (external) objects in A+ code. We collected the object oriented concepts and investigated the required concepts for the A+ language to conveniently handle objects. Based on the result of our investigation, we have extended the A+ language with new language elements. For these language elements new symbols are added and they follow one of the main characteristic of the A+ language which is the right-to-left evaluation order.

Each of the new language elements represents basic operations which are needed to handle various tasks on objects. These operations are the following:

- Accessing methods, variables and properties.
- Modifying variables and properties.
- Type casting.
- Using indexer properties.

We present the mechanism behind each operation and also provide examples on how to use them. With these basic operations, most of the .NET classes can be accessed and used from A+ scripts without writing any additional wrapper methods.

The presented language extension is runtime agnostic and with it's help it would be possible to connect to other object oriented runtimes also. A major result of the extension is that there is now a more convenient way working with objects which are from the .NET world. Also the extension paves the way to add full object oriented support for the A+ language.

References

- [1] P. Gál and Á. Kiss. Implementation of an A+ interpreter for .NET, In *Proceedings of the 7th International Conference on Software Paradigm Trends 2012*, pages 297-302, SciTePress.
- [2] Morgan Stanley, A+ Language Reference, <http://www.aplusdev.org/Documentation/>

Common Open Telemedicine Hub and Interface Standard Recommendation

Ábel Garai, István Péntek

The presented paper demonstrates the gaps of health systems interoperability with solution recommendation. This article shows the conclusions of the research on clinical integration of eHealth smart device technology. The aim is to establish the empirical basis for the general IoT-Clinical Systems interconnection. The emerging sensor-based smart devices collect bulk data. The technical solutions for building a bridge between the classical clinical information systems and the eHealth smart devices is still missing. Sensor-based smart devices collect and transmit continuous data series. Clinical information systems transmit and store static data with reference to patients. The data representation, structure, methodic and rationale is different by the hospital information systems and by the eHealth smart devices. Present IoT trends show that within a decade multi-billion smart devices will be communication continuously multilaterally. The issue of interoperability among clinical systems and eHealth smart device technology is unresolved. There is a need for the flawless technical and methodical interoperability among the classical clinical information systems, the industrial telemedicine instruments and the eHealth smart appliances. For the study a hospital information system installed in sixty hospitals within Europe and serving around forty thousand users was analyzed. Furthermore, a medical spirometer was installed and connected through HL7-based interface to the hospital information system. A heart rate monitor smart device was also procured for the research. There are international standards and recommendations for clinical systems interoperability. However, the integration and unification of these standards and recommendations is an issue to be solved. Furthermore the interconnection of the eHealth smart devices with the classical hospital information system landscape is a significant challenge. As the eHealth smart device users number will reach billions within the near future, the collected healthcare information have remarkable value. Recent big data analytical capacities combined with the incoming volumes of personal health data open new horizons of forecasting both at personal and at community level, including epidemic control. The bandwidth capacity of one GSM network was also analyzed in order to serve as an empirical basis whether the telecommunication infrastructure is prepared for the challenges posed by the massive data-exchange of the IoT. In this paper the leading international clinical interoperability standards, inter alia Snomed and HL7 are analyzed and demonstrated. The challenges blocking the proliferation of the HL7 v3 standard is also presented within this paper. The research takes place at the Department of Pulmonology of a Pediatric Clinic. A spirometer is connected to the clinical information system during the presented research. The clinical information system is accessed with tablets through local clinical wlan by the medical staff. The research demonstrates, how the mobile spirometer and the also mobile clinical information system GUI works together, enabling further application areas of this system and architectural landscape, like in mobile ambulance or deployment in remote or rural areas. The presented paper demonstrates the different HL7 v2.x and HL7 v3 standards. Notwithstanding that the latest HL7 v3 standard has a wider capacity range, the HL7 v2.x standard is going to dominate the interoperability platforms among the clinical information systems. Beside clinical systems operability, as mentioned earlier, the integration of the eHealth smart device technology within the classical medical system landscape is highly desirable. Therefore the Common Open Telemedicine Interoperability Hub and interface standard recommendation is presented within this paper. The Common Open Telemedicine Interoperability Hub (COTIH) establishes the link for valid information exchange among the classical medical systems, industrial telemedicine instrument, eHealth smart devices and adaptive health-services. The COTIH relies upon the basis of the international HL7 standard, and with thorough extension it provides with interconnection capabilities

reaching the eHealth smart devices. The paper presents the interoperability modalities concretizing the contextual background. The cloud architecture lies underneath the COTIH. The cloud architecture brings significant benefits, however leads to legal obstacles, issues of confidence and of personal perceptions. The article concludes that notwithstanding that the COTIH enables stable semantic interoperability between the classical medical information systems and the eHealth smart device technology, there is further room for improvement regarding process interoperability, which is also necessary for the flawless multilateral communication of health-related systems and appliances.

Acknowledgements

This work was supported by Attila Adamkó, Department of Information Technology, University of Debrecen.

References

- [1] Howard, E., Bhardwaj, V. HL7 for BizTalk. Springer/Apress, New York, 2014.
- [2] Fong, B., Fong, A. C. M., Li, C. K. Telemedicine Technologies. Wiley, Chichester, 2011.
- [3] Sicilia, M. A., Balazote, P. Interoperability in Healthcare Information Systems. IGI Global, Herchey, 2013.
- [4] Somanm, A.K. Cloud-based Solutions for Healthcare IT. CRC Press, Science Publishers, Enfield, 2011.
- [5] Thoma, W. HL7 konkret - Analyse und Beschreibung der HL7 Kommunikation im Kontext eines Krankenhauses. Akademiker Verlag, Saarbrücken, 2014.
- [6] Garai, Á., Péntek, I. Adaptive Services with Cloud Architecture for Telemedicine. Proceedings of the 6th IEEE Conference on Cognitive Inforcommunications; October 19, 2015; Győr, pp. 369-374, doi: 10.1109/CogInfoCom.2015.7390621.

Supervising Adaptive Educational Games

László Gazdi, Dorottya Bodolai, Dr. Bertalan Forstner Ph.D., Luca Szegletes

Recent years, the number of children with special needs have risen. Researches say that approximately 5 to 10 percent of the children have some kind of learning disability. They are not able to adapt the traditional teaching methods, and because of this, they cannot take part in the traditional education system. Their special needs and changed perception ability require highly qualified experts and special teaching methods. If they get this necessary treatment, they can reach as high results as their healthy mates. Researches say that children with special needs can communicate better, if digital devices are included to the communication. It is the same at their learning. The recent years educational pieces of software were developed for students with dyscalculia, dyslexia, dysgraphia or ADHD (Attention Deficit Hyperactivity Disorder). These tools are giving a great opportunity for individual practicing, which is very important for these children. Nevertheless, they have a problem. We cannot expect from a child to be able to set the proper task and difficulty for himself. These values have a huge effect on the learning efficiency. If they are chosen badly, the student quickly can get bored or frustrated and demotivated about the learning.

In this paper, we present our universal solution to this problem. In our system, which runs on mobile devices, there is a framework, which can measure the mental state of the student with the attached biofeedback devices. Based on these measurements, it is able to make recommendations to change the difficulty of the gameplay, or to give a reward to the player. With these changes, it can affect the mental state of the student, in order to keep him in the proper state. The framework contains an algorithm, which is responsible for the calculation of the mental state and the mental workload.

Our system makes it possible for the teacher to set up the training set manually, and make the machine learning process supervised. The framework contains a function, where the supervisor can send feedback to the system to train the classification algorithm. The teacher can make the recommendation based on her personal experience or the physiological data of the student, which is presented to her by the framework. Our system is designed not only for one student. It also helps the teacher to supervise a whole class efficiently. She is able to monitor and manipulate the gameplay of many students at the same time.

An Automatic Way to Calculate Software Process Metrics of GitHub Projects With the Aid of Graph Databases

Péter Gyimesi

Bug prediction is a popular research area nowadays. There are many great studies about how to characterize software bugs and how to recognize bug prone code parts. The characterization [1] of such bugs can be done with classic product metrics, with software process metrics or with some metrics of a different nature like textual similarity. Product metrics are extracted from the structure of the source code, for example: lines of code, cyclomatic complexity, etc. Software process metrics are computed from developer activities. The most common ones are based on the number of previous modifications, number of different contributors, number of modified lines and the time of the modifications.

All of these metrics can be computed on different granularity levels (file, class, method). Product metrics are mainly calculated on file and class level, but with the continuously improving technology the method level is becoming more frequent. There are many great tools – some of them are free – that can produce these metrics for projects of different programming languages. Previous researches [2] have shown that software process metrics are generally better bug predictors than product metrics, however, tools that can compute these metrics are not widespread. It can be due to the difficulties of storing and processing of the historical information. Another problem with process metrics is that to compute them on class or method level every version of the source code has to be analyzed and the modified methods and classes have to be extracted, which is a challenging task.

In the last few years, the popularity of graph databases increased due to the improving technologies behind them. The historical data of software can also be represented as a graph, so studies [3] started to examine the use of these graph databases in software quality too, especially in the calculation of software process metrics.

Our goal is to present an automated way of calculating software process metrics with the use of graph databases. We chose GitHub as data source, because it has an open API to access the history of many open-source Java projects. For the database we selected Neo4j and to analyze the source code and extract the source code elements we used the SourceMeter static analyzer tool.

For this research, we collected the process metrics from the literature and we reproduced the calculation of these metrics in cypher query language. With this language process metrics are easy to formulate and these queries can be processed by Neo4j. For a given GitHub project version it can compute the process metrics on file, class and method level. The list of these metrics is easy to extend due to the modularity of the created framework.

With this tool we analyzed almost 18 000 versions of 5 Java projects from GitHub. These projects have a total of 28 release versions that are selected with six-months-long intervals. We computed the process metrics of these projects and we built a bug database with the use of our previous researches. For each release version it contains the source code element at file, class and method level with the corresponding process metrics and bug numbers.

References

- [1] Radjenović, Danijel, et al. "Software fault prediction metrics: A systematic literature review.", *Information and Software Technology* 55.8 (2013): 1397-1418.
- [2] Rahman, Foyzur, and Premkumar Devanbu. "How, and why, process metrics are better.", *Proceedings of the 2013 International Conference on Software Engineering*. IEEE Press, 2013.

- [3] Bhattacharya, Pamela, et al. "Graph-based analysis and prediction for software evolution.", *Proceedings of the 34th International Conference on Software Engineering*. IEEE Press, 2012.

The structure of pairing strategies for k -in-a-row type games

Lajos Győrfy, András London, Géza Makay

The positional game k -in-a-row is played by two players on the infinite (chess)board. In the classical version of the game the players alternately put their own marks to previously unmarked squares, and whoever reaches a *winning set* (k -consecutive squares horizontally, vertically or diagonally) first of his own marks, wins. In the Maker-Breaker game Maker is the one who tries to occupy a winning set while Breaker only tries to prevent Maker's win. For different values of k , there are several winning strategies either for Maker or for Breaker. One group of those are *pairing* (*paving*) *strategies*.

A pairing strategy generally means that the possible moves of a game are paired up; if one player plays one, the other player plays its pair. A winning pairing strategy of Breaker in the k -in-a-row type games is such a pairing of the squares of the board that each winning set contains at least one pair. Using the above mentioned pairing strategy, Breaker gets at least one square from each winning set, therefore wins the game. We showed that Breaker can have a winning pairing strategy for the k -in-a-row game, if $k \geq 9$.

In the case of $k = 9$ there are only highly symmetric (8- or 16-toric) pairings of the board, we list all (194543) 8-toric pairings. Among those 194543 different pairings we can define alternating paths and cycles which give us a natural link between two pairings: Two pairings are neighboring if we can get them from each other by alternating only one path or cycle.

We analyze the network of pairings which is a huge and sparse graph (194543 nodes and 532107 edges). It is triangle free and has 14 components. One of those components is a giant component containing 194333 nodes and there are some components which forms the net of a 4D cube.

We investigate another similar game by considering the k -in-a-row game on the hexagonal board. For that game Breaker has winning pairing strategies if $k \geq 7$. For $k = 7$ we list 26 different 6-toric pairings – and these are the only ones – which gives us a similar but smaller graph as before.

At the end we highlight some possible extension to higher dimensions and the issues arising from those.

References

- [1] J. Beck, Combinatorial Games, Tic-Tac-Toe Theory, Cambridge University Press 2008.
- [2] A. Csernenszky, R. Martin and A. Pluhár, On the Complexity of Chooser-Picker Positional Games, *Integers* 11 (2011).
- [3] L. Győrfy, G. Makay, A. Pluhár, Pairing strategies for the 9-in-a-row game, submitted (2016+)
- [4] L. Győrfy, A. Pluhár, Generalized pairing strategies, a bridge from pairing strategies to colorings, submitted (2016+)

Maximization Problems for the Independent Cascade Model

László Hajdu, Miklós Krész, László Tóth

While the research of infection models for networks (directed graphs) is a classic field, its business application has only become widespread in recent years. The definition of the Independent Cascade Model in [1] is the most important breakthrough from a computational theory point of view. The infection spreads the following way: we choose an initial infected vertex set, and only those vertices spread the infection after this that became infected in the previous iteration. Every edge starting from an infected vertex has a single chance for infecting using its own infection probability. Kempe et al. [1] defined the influence-maximization problem, where we are looking for the vertex set for which the expected value of the infected vertices is the highest. This optimization problem is NP-complete, but Kempe et al. proved in [1] that applying a greedy algorithm gives a guaranteed precision that results in good quality solutions in practice. Bóta et al. [2] extended the model and introduced the Generalized Independent Cascade Model. The initial state of this model does not consist of *infected* and *uninfected* vertices, but introduces an *a priori* probability of infection for each vertex. By the end of the infection process, each vertex receives an *a posteriori* infection value. Because the infection process above is #P-complete, [2] introduces a number of approximation algorithms. However, the influence-maximization problem has not been defined for this more general case. In this talk we define two infection-maximization problems connected to the generalized model above. We also present solution methods based on greedy heuristics, and compare their results with each other. We prove that a guaranteed approximation precision can be achieved for these greedy algorithms. To guarantee efficiency from a practical point of view, we decrease the search space for the greedy methods using different approaches (e.g. selection based on vertex degree). We also present a new method that chooses the vertices during the infection process by applying metrics based on the communities (dense subgraphs) of the graph.

References

- [1] D. Kempe, J. Kleinberg, E. Tardos Maximizing the Spread of Influence through a Social Network, *Proceeding of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, 137-146 (2003)
- [2] A. Bóta, M. Krész, A. Pluhár Approximations of the Generalized Cascade Model, *Acta Cybern.* 21, 37-51 (2013)

Eliminating code coverage differences from large-scale programs

Ferenc Horváth

During software development special activities are done to keep the quality of the software while the requirements and the code are constantly changing. This includes, white-box test design, massive regression testing, selective retesting, efficient fault detection and localization, as well as maintaining the efficiency and effectiveness of the test assets on a long term [1]. These activities are usually based on *code coverage*, a test completeness measure, therefore, code coverage measurement is an important element both in industrial practice and academic research. Obviously, inaccuracies of a code coverage tool sometimes do not matter that much but in certain situations they can lead to serious confusion. For Java, the prevalent approach to code coverage measurement is to use bytecode instrumentation due to its various benefits over source code instrumentation. However, there can be differences in the list of covered items reported by the two approaches, and these differences have influence on the information derived from them. Tengeri *et al.* [2] investigated this two types of code coverage measurement on Java systems: they analyzed the results of method level coverage measurements on Java programs and found that there are many deviations in the raw coverage results due to various technical and conceptual differences of the instrumentation methods. Similar studies exist in relation to branches and statements [3], where the authors investigated how the differences can impact further activities.

In this paper, I extend the work of Tengeri *et al.* [2] and present the results of an empirical study conducted on eight large-scale programs, concentrating on how the tools can be configured and the results be filtered so that the causes are eliminated and the differences in the coverage results are alleviated as much as possible. In addition, I present my experiences on how big difference remains after eliminating tool-specific differences, which can be possibly attributed to the differences in the fundamental approach, that is, bytecode vs. source code instrumentation. The results indicate that in the programs with the biggest overall coverage difference, about 10% of the methods are falsely reported as not covered, and that when all factors including not recognized and fewer times covered cases are taken into account, the difference is much more emphasized. On average, 20% of all methods of the eight programs have false coverage data, when investigated more closely.

References

- [1] L. S. Pinto, S. Sinha, and A. Orso, "Understanding myths and realities of test-suite evolution", in *Proceedings of the ACM SIGSOFT 20th International Symposium on the Foundations of Software Engineering*. ACM, 2012, pp. 33:1–33:11.
- [2] D. Tengeri, F. Horváth, Á. Beszédes, T. Gergely, and T. Gyimóthy, "Negative effects of bytecode instrumentation on java source code coverage", in *Proceedings of the 23rd IEEE International Conference on Software Analysis, Evolution, and Reengineering (SANER 2016)*, Mar. 2016, pp. 225–235.
- [3] N. Li, X. Meng, J. Offutt, and L. Deng, "Is bytecode instrumentation as good as source code instrumentation: An empirical study with industrial tools (experience report)," in *Software Reliability Engineering (ISSRE), 2013 IEEE 24th International Symposium on*, Nov 2013, pp. 380–389.

Multiple Intelligence Learning

Laura Horváth, Krisztián Pomázi, Bertalan Forstner, Luca Szegletes

Recently, the method to determine a person's intelligence and intellectual performance has been the usage of the intelligence quotient. There are many factors that this method does not take into consideration. Certain parameters related to the person, such as age, mental state, learning disabilities are not dealt with. Gardner's research points out that there may be more efficient ways to assess cognitive capabilities as it cannot be decided by regarding only one ability. Following this concept, our AdaptEd project approaches the issue by using biofeedback devices and automatized learning in order to enhance a person's learning process by stimulating different cognitive abilities and then adapting to them.

There is an increasing need for different teaching methods and a demand for schools which educate based on children's individual competences. We can see from the rising numbers of alternative schools, and the growing need for them, that both parents and decision-makers become aware of the need to teach each child according to her skills. It is important to obtain information about a child's cognitive performance before making a comparison to others in regard to intelligence. It is possible that with different teaching methods, children with disabilities could get similar results to their classmates.

It is necessary to analyse a person's mental workload during certain tasks and establish a model to be able to estimate her mental effort later. Using biofeedback data from sensors, such as heart rate monitoring and electroencephalography can be useful for establishing the model. If the person is required to do exercises of different difficulty level while this kind of sensor data is constantly being monitored, it is possible to separate different stages of mental workload. Given the fact that the levels of performance during these are personal characteristics, it is possible to adjust the model and thus, consider differences between people with different cognitive capabilities.

With measuring the different fields of intelligence, we are able to get a clear picture about a child's strengths and weaknesses, and provide for her the education that best fits her abilities. For creating this individually-optimised education, the use of today's technology, the use of tablets and smartphones is required.

It is also an important step to determine to what extent and how exactly should the sensor data be used for the adjustment to the individual and for further estimation, what kind of processing and mathematical calculations can be considered. Our hypothesis is that with the use of biofeedback sensors and the AdaptEd framework we are able to test the multiple intelligence fields and make numerical measurements possible.

Our solution assembles, processes and stores biofeedback data in order to create a software environment where the mental workload can be estimated after a previous assessment of the person's abilities, then continuously estimates the workload and can adjust itself in difficulty to the person. The ideal result is to successfully find out if the given task matches the subject's cognitive abilities and then increase or reduce the difficulty.

We decided to build our solution on top of a framework, which integrates physiological signals into educational gaming on a mobile device. The software package has four major interfaces: for the biofeedback sensors, for the different (game) software components, one towards a supervisor machine and for the backend side. The application gathers data from the sensors, and uses a background process to upload them to the server.

In order to train the classification algorithm and also to measure the separate intelligence fields, we employ custom made cognitive games. Such games are N-back with symbols, numbers, characters or sounds, anagram guessing game, visual memory, or 3D abstraction games, which we implemented during our research. With the correct parametrization and setup of these games, we can infer on the intelligence and skills of the user. This wide variety of cog-

nitive games enables us to appropriately measure the following intelligence fields: musical-rhythmic, logical-mathematical, and verbal-linguistic. All of the games are available for mobile use as well.

To determine the actual mental effort we chose to use linear SVM (Support Vector Machine) for classification based on real-time sensor values measuring inter-beat intervals for heart rate and/or measured EEG values. First, a model needs to be established so the algorithm can learn what kind of values exactly can be linked to different stages of mental effort, meaning, a training period needs to precede the mental workload estimation. The training period consists of different phases, each one of them giving a certain exercise or activity that the person in question has to do or solve. To achieve a better set of training data, the exercises are of well separable difficulty.

The software environment can be used with various learning games with different kinds of tasks. For the individual and private experience, a feedback about the smartphone or tablet usage is required, and biofeedback sensors are the best way to get these results. By measuring a child's brain waves and heart signals, we can get a clear knowledge about her mental workload and mental state, and optimize our educational tools for better results and better experience.

We created the software and hardware environment for our solution, in which we focus on signal processing and estimation of the mental workload, as we consider these the biggest challenges. In the future we plan to fine-tune our model with further measurements and analysis of the data. Our objective is to be able help children with or without learning disabilities have better learning experience and better results in the long term.

The Optimization of a Symbolic Execution Engine for Detecting Runtime Errors

István Kádár

Nowadays, producing great, reliable and robust software systems is quite a big challenge in software engineering. About 40% of the total development costs go for testing [1] and maintenance activities, moreover, bug fixing of the system also consumes a considerable amount of resources [2].

The symbolic execution engine called RTEHunter developed at the Department of Software Engineering at the University of Szeged supports this phase of the software engineering lifecycle by detecting runtime errors (such as null pointer dereference, bad array indexing, division by zero) in Java programs without actually running the program in real-life environment.

During its execution, every program performs operations on the input data in a defined order. Symbolic execution [3] is based on the idea that the program is operated on symbolic variables instead of specific input data, and the output will be a function of these symbolic variables. A symbolic variable is a set of the possible values of a concrete variable in the program, thus a symbolic state is a set of concrete program states. When the execution of the program reaches a branching condition containing a symbolic variable, the condition cannot be evaluated and the execution continues on both branches. The execution paths created this way compose a tree called symbolic execution tree. At each branching point both the affected logical expression and its negation are accumulated on the true and false branches, thus all of the execution paths will be linked to a unique formula over the symbolic variables.

RTEHunter detects runtime issues by traversing the symbolic execution tree and if a certain condition is fulfilled the engine reports an issue. However, the number of execution paths increases exponentially with the number of branching points thus the exploration of the whole symbolic execution tree is impossible in practice. To overcome this problem the symbolic execution engines set up different kinds of constraints over the tree. E.g. the number of symbolic states, the depth of the execution tree, or the time consumption is limited. In RTEHunter the depth of the symbolic execution tree (means symbolic state depth) and the number of states can be adjusted and the strategy of the tree traversal can also be selected.

Our goal in this work is to find the optimal parametrization of RTEHunter in terms of maximum number of states, maximum depth of the symbolic execution tree and search strategy in order to find more runtime issues in less time. The maximum depth limits the height of the tree, the maximum number of states defines its width, while the search strategy determines in which order the states in this limited size tree should be traversed.

The results show that by limiting only the depth which results a wide and flat-shaped tree the number of detected runtime errors increases exponentially in depth. However, the time consumption also increases exponentially indicating that such a flat tree configuration is not efficient in terms of the number of detected issues per time unit.

Better results can be achieved by a narrow but deeper tree. Limiting the tree by around 300 depth and 3000 states the number of detected runtime issues increased by 55% within the same time frame compared to the flat-shaped tree turning out the best configuration considering the shape of the tree.

We also developed a search strategy that improves the performance. This search strategy uses a heuristic which is based on the number of null values in a state, and directs the search towards those states where there are more null values, increasing the chance to find null pointer dereference issues. The number of the detected null pointer dereferences increased by 13% on average using this heuristic approach compared to the general breadth-first search strategy.

References

- [1] Pressman, Roger S., *Software Engineering: A Practitioner's Approach*, McGraw-Hill Higher Education, 2001
- [2] Tasse, G., *The Economic Impacts of Inadequate Infrastructure for Software Testing*, National Institute of Standards and Technology, 2002.
- [3] King, James C., *Symbolic Execution and Program Testing*, Communications of the ACM, vol. 19, no. 7, pp. 385-394, 1976.

Image processing based automatic pupillometry on infrared videos

György Kalmár, László G. Nyúl

Pupillometry is a non-invasive technique that can be used to objectively characterize pathophysiological changes involving the pupillary reflex. It has long been utilized in humans to measure the pupil diameter and its dynamic changes. In animal models of human disorders, pupillometry derived reflex metrics could potentially be used to monitor the extent of disease and response to therapy [1]. Specially designed computer algorithms provide faster, more reliable and reproducible solutions to the measuring process. These methods use a priori information about the shape and color of the pupil [2, 3].

Our research focuses on measuring the dynamics and diameter of the pupil of rats from videos recorded with a modified digital camera under infrared (IR) illumination. In these experiments the left eye of the animal is being imaged while the right eye is stimulated with a visible light impulse. The main goal is to analyze the relative size change of the pupil and the velocity of the contraction and recovery. During the recordings, the rats are breathing and moving slightly that changes their distance from the camera. Since the lens is very close to the animal, the video is distorted by scattering movements and also significant blur due to the eye getting out of focus. Essential pre-processing of the videos includes motion compensation and contrast enhancement. To avoid scale related errors, eye segmentation is performed on each frame to give a reference measure to later phases. Then the centre of the pupil is found by an energy-based ray-tracing voting system with Least-Square Estimation. In the last phase, the diameter of the pupil is determined in each frame, and at the end, the required features from the video are extracted. We developed a new, robust method that can reliably measure the size of the pupil under the mentioned difficult circumstances, compare our results with measurements obtained using manual annotation and discuss the reliability and accuracy of our method.

Acknowledgements

The authors are thankful to Prof. Dr. Gyöngyi Horváth for providing the video data sets.

References

- [1] R. Z. Hussain and S. C. Hopkins and E. M. Frohman and T. N. Eagar and P. C. Cravens and B. M. Greenberg and S. Vernino and O. Stuve, *Direct and consensual murine pupillary reflex metrics: Establishing normative values*, Autonomic Neuroscience, vol. 151, no. 2, pp. 164-167, 2009.
- [2] D. Iacoviello and M. Lucchetti, *Parametric characterization of the form of the human pupil from blurred noisy images*, Computer Methods and Programs in Biomedicine, vol. 77, no. 1, 2005.
- [3] A. De Santis and D. Iacoviello, *Optimal segmentation of pupillometric images for estimating pupil shape parameters*, Computer Methods and Programs in Biomedicine, vol. 84, no. 2-3, pp. 174-187, 2006.

Quantitative Assessment of Retinal Layer Distortions and Subretinal Fluid in SD-OCT Images

Melinda Katona, László G. Nyúl

A modern tool for age-related macular degeneration (AMD) investigation is Optical Coherence Tomograph (OCT) that can produce high resolution cross-sectional images about retinal layers. AMD is one of the most frequent reasons of blindness in economically advanced countries. AMD means the degeneration of macula which is responsible for central vision. Since AMD affects only this area of the retina, unattended patients lose their fine shape-, and face recognition, reading ability and central vision [1].

The constant growing of AMD patient population is more and more challenging. Consequently, there is a need for more precise measurements and search for new OCT features for earlier and more efficient treatment. Quantitative assessment of retinal features in course of AMD is not yet fully explored. Existing OCT systems are partially suited to monitoring the progress of the disease. There are some typical symptoms of AMD that can be sensitive retreatment criteria. Digital image processing can help the quantitative assessment of the OCT features of AMD by providing automatic tools to detect abnormalities and to describe, by objective metrics, the current state and longitudinal changes during disease evolution and treatment.

In this paper, we deal with automatic localization of subretinal fluid areas and also analyze retinal layers, since layer information can help localizing fluid regions. Image quality is improved by a noise filtering and contrast enhancement method using fuzzy operator [2]. This operator can highlight major retinal layers so we analyzed vertical profiles of the filtered image and high intensity steps in pixel density are assumed to correspond to change of tissues. We present an algorithm that automatically delineates seven retinal layers, successfully localizes subretinal fluid regions, and computes their extent and volume. We show our result using a set of SD-OCT images. The quantitative information can also be visualized in anatomical context for visual assessment. Our results were verified by ophthalmologists and they found the segmentation, quantification and also the visualization technique useful.

Acknowledgements

This research was supported by the European Union and the State of Hungary, co-financed by the European Social Fund in the framework of TÁMOP-4.2.2.D-15/1/KONV-2015-0024 'National Excellence Program'. The authors are grateful to Dr. Rózsa Dégi and Dr. Attila Kovács for providing the OCT data sets and their medical expertise. The authors also thank Dr. József Dombi for suggesting the use of fuzzy operators for image preprocessing.

References

- [1] M. R. Hee, C. R. Baumal, C. A. Puliauto, J. S. Duker, E. Reichel, J. R. Wilkins, J. G. Coker, J. S. Schuman, E. A. Swanson, and J. G. Fujimoto, *Optical coherence tomography of age-related macular degeneration and choroidal neovascularization*, Ophthalmology, vol. 103, no. 8, pp. 1260-1270, 1996.
- [2] J. Dombi, *Modalities*, Eurofuse 2011: Workshop on Fuzzy Methods for Knowledge-Based Systems, pp. 53-65, 2012.

Height Measurement of Cells Using DIC Microscopy

Krisztian Koós, Begüm Peksel, Lóránd Kelemen

Development of fluorescent probes and proteins increased the use of light microscopy notably by enabling the visualization of subcellular components, location and dynamics of biomolecules. However, it is not always feasible to label the cells as it may be phototoxic or perturb the function. On the contrary, label-free microscopy techniques enable the work with live cells without perturbation and the evaluation of morphological differences which in turn can offer useful information for the high-throughput assays. In this work we use one of the most popular label-free techniques, Differential Interference Contrast (DIC) microscopy to estimate the height of cells.

DIC images show detailed information about the optical path length (OPL) differences in the sample. DIC images are visually similar to a gradient image. The mathematical description of the nonlinear image formation model is a recent research topic [1]. In their earlier work [2], the authors proposed a DIC reconstruction algorithm to retrieve an image where the values are proportional to the OPL (or implicitly the phase) of the sample. Although the reconstructed images are capable of describing cellular morphology and to a certain extent turn DIC to a quantitative technique, the actual OPL has to be computed from the input DIC image and the microscope calibration properties. The retrieved phase has already been compared to synthetic data or objects with known properties [3]. Here we propose a computational method to measure the approximate height of cells after microscope calibration. The method starts with calibrating the microscope by setting the Koehler illumination and ensuring that the output image utilizes the full dynamic range of the camera. Then the DIC image of a calibration object with known dimensions and refractive index has to be reconstructed. This allows the mapping of a unit change of image intensity to OPL change assuming a linear formation model. The calculated ratio can be used to determine the height of further samples where the refractive index of the surrounding medium is known. The described method converts DIC to a quantitative technique. The method's precision is tested on a special calibration sample.

Acknowledgements

K. K. acknowledges the financial support from the National Brain Research Programme (MTA-SE-NAP B-BIOMAG). We thank Peter Horvath for his useful discussions and László Vígh for his help.

References

- [1] Mehta, S. B. & Sheppard, C. J. Partially coherent image formation in differential interference contrast (DIC) microscope. *Opt. Express*, 16, 19462-19479 (2008).
- [2] Koos, K., Molnár, J., Kelemen, L., Tamás, G., & Horvath, P. DIC image reconstruction using an energy minimization framework to visualize optical path length distribution. Submitted to *Nature Scientific Reports*.
- [3] King, S. V., Libertun, A., Piestun, R., Cogswell, C. J., & Preza, C. Quantitative phase microscopy through differential interference imaging. *Journal of Biomedical Optics*, 13(2), 024020 (2008)

Robust multiple hypothesis based enumeration in complex traffic scenarios

Levente Lipták, László Kundra, Dr. Kristóf Csorba

The measurement of the traffic load on the road network is essential for calibrating the traffic lights and designing future developments. One of the first published solutions is dated back to 1937, in which a strip through the road was used to count vehicles, with an electronic device attached at the end. Nowadays this work is still done by so-called enumerators, human traffic counters. This is due to the very strict precision requirements, and the handling of the complicated traffic situations. Our aim is to ease their work with a high accuracy vehicle counter software[1][2]. The videos are produced by cameras placed on street furniture, and processed with Open Computer Vision Library (OpenCV) providing statistics of the vehicle types and their motion path. The application has to handle varying weather conditions, the continuously changing lights, and shadows. The cameras are deployed in new junctions every day, lead to different viewpoints at different junctions, and diverse traffic situations. The ideal camera position would be high above the center of the junction, in top view, avoiding any overlapping. Their positions are limited to the reachable locations, which are usually not ideal for image processing, and lead to overlapped vehicles. As the industry still employs enumerators in many cases, the recorded videos have often low quality which is sufficient for humans, but contain lots of noise for automatic processing. Because of this our system has to be prepared for very low quality input as well.

The videos are processed with multiple methods to provide the best results. Detections of slow or stopping vehicles are poor using only motion based methods due to the background model contamination effect merging vehicle images into the background model. However, this is a simple task for methods based on easily recognizable unique shapes (Maximally Stable Extremal Regions) or corner points (Shi-Tomasi Corner Detector). In the case of a fast vehicle their performances are reversed. Fusing together information extracted from the image simultaneously is not a simple task, because detections are described by different data types, and also have dissimilar noise to be eliminated. Though the produced set of information is complex, it is necessary to define the criteria of similarity to be able to pair identical objects on consecutive frames. Furthermore the system should be able to handle a situation containing overlapping vehicles moving on the same or on different paths.

To reach the highest accuracy, multiple hypotheses are created in the more complicated situations, increasing the probability that they include the one representing the correct situation. More of them could be useful, if some parts of the movement curves of multiple vehicles are joined. However, some of them are useless, and they have to be filtered out in possession of more information. Erroneous hypotheses could be in the result set too (for example pedestrians and wind-blown objects), which also have to be filtered. This paper is presenting the conclusion of 2 years of research, aiming at main architecture, hypothesis building, and filtering of the duplicates and improbable hypotheses based on a physical model.

References

- [1] V. Kastrinaki, M. Zervakis, K. Kalaitzakis: A survey of video processing techniques for traffic applications, In Image and Vision Computing, Vol. 21, Elsevier, 2003. pp. 359-381.
- [2] Norbert Buch, Sergio A. Velastin, James Orwell: A Review of Computer Vision Techniques for the Analysis of Urban Traffic IEEE Transactions on Intelligent Transportation Systems (Volume:12 , Issue: 3), 2011, pp 920 - 939

Unit Testing and Friends in C++

Gábor Márton, Zoltán Porkoláb

Unit testing of object-oriented code requires two fundamental artifacts: replacing some of the original code with some kind of test doubles (aka dependency injection), and providing extra access to these private dependencies. In most cases we cannot apply link-time or preprocessor based dependency injection, therefore a typical testable unit uses additional interfaces or template type parameters for pure testing purposes. Thus a unit which we developed without testing in mind might have to be changed intrusively for testing. These changes can take us far away from the original natural design of the unit. As a result, in C++, test code is often badly interwoven with the unit we want to test. We've seen such interleaving in the software world before: exceptions and aspect oriented programming were both invented to try to eliminate this kind of interleaving of independent aspects.

In this paper we demonstrate the above problems using code examples to show the decay of original intentions. We discuss current possible directions to solve the dependency injection problem: e.g. using compile-time seams [1]. Then, we overview the current (partial) solutions for adding extra access for private data, like using friends, the Attorney-Client idiom [2], or exploiting privates using explicit template instantiations.

Regarding to extra access of private data, we present how a minimal language extension would help and make the testing code less intrusive. Open non-intrusive friend declarations could provide a good method to separate the test related code from the original intentions. We implemented a proof-of-concept solution, based on LLVM/Clang to show that such constructs can be established with a minimal syntactical and compilation overhead.

References

- [1] M. Rüegg, *Refactoring Towards Seams in C++*, ACCU OVERLOAD, #108, ISSN 1354-3172, April 2012.
- [2] A. R. Bolton, *Friendship and the Attorney-Client Idiom* Dr.Dobbs's Journal, January 01, 2006.

Revision of local search methods in the GLOBAL optimization algorithm

Abigél Mester, Balázs Bánhelyi

There are many suitable global optimization methods to find the minimum of a nonlinear objective function. In our presentation we improve the GLOBAL optimization method which is developed byin the Institute of Informatics, University of Szeged. GLOBAL was used successfully for the solution of very complex optimization problems.

GLOBAL is a stochastic technique that is a sophisticated composition of sampling, clustering, and local search. It compares well with other global optimization software for the solution of the low dimensional black-box-type problems (when only the objective function is available, while the derivatives should be approximated). It is usually very successful on problems where the relative size of the region of attraction of the global minimizer is not negligible.

The performance of the GLOBAL depends mainly on the applied local search method efficiency. In our presentation we try to improve the local search methods. First of all, we changed the original Unirandi random walk technique, so the direction search and the line search method became independent modules. After this we recreated the Rosenbrock method with the same modular structure. Then we have developed two new different line search methods based on polynomial interpolation schemes. In the presentation we show the efficiency of these method combinations, and the advantages of the modular structure.

Multi-layer phase field model for selective object extraction

Csaba Molnár, József Molnár

Selective object segmentation is a common problem of image processing. This is one of the first steps of most image analysis applications, therefore its accuracy is crucial for the later calculations. The criteria of selection varies from the simplest ones such as the intensity or texture properties of objects to more complex ones like object shape descriptors. Earlier methods use object templates as reference shapes. This approach has limited flexibility and not efficient for unknown number of instances. Variational models commonly used to describe shapes without model templates. A special type of these models called higher-order active contours, are based on contour representation of regions and use interactions between contour points to describe regular shapes [1]. Numerous applications in biology (e.g. cell nuclei detection on fluorescent microscopy images) and physical sciences (e.g. nanoparticle delineation in transmission electron microscopy) objects can touch or even overlap with each other [2]. In this work, we present a multi-layer phase field segmentational model that can extract a family of shapes with predefined size and shape characteristics such as ratio of area and perimeter. With multiple phase-field layers we can represent overlapping objects. Using a common phenomenon in microscopy that is the measured intensities are additive if objects located on the top of each other. We combine the size and shape selective prior model with this additive data model which can handle overlapping parts if they have the above properties. We tested the model on synthetic and on real microscopy images.

Acknowledgements

We acknowledge the financial support from the National Brain Research Programme (MTA-SE-NAP B-BIOMAG).

References

- [1] Horvath, P., *et al.* A higher-order active contour model of a 'gas of circles' and its application to tree crown extraction, *Pattern Recognit*, **42**(5), 699–709 (2009).
- [2] Molnar, C., *et al.* A New Model for the Segmentation of Multiple, Overlapping, Near-Circular Objects, In *DICTA 2015: Proceedings of International Conference on Digital Image Computing: Techniques and Applications*, 1–5, Adelaide, Australia, IEEE (2015).

Skyline extraction by artificial neural networks for orientation in mountainous environment

Balázs Nagy

Skyline or horizon line is a salient feature in a mountainous scenery and it can be applied for orientation. Skyline extraction from an image is a complex task because plenty of circumstances can vary. In many problems the accurate azimuth angle is essential e.g.: visual localization, annotation of geotagged photos or augmented reality applications. Mobile devices are equipped with various sensors from among which the GPS sensor determines the position, while the fusion of the magnetometer, accelerometer and the gyroscope sensors measures the translation and rotation of the observer in 3D space. Despite the frequent calibration, the magnetometer may be biased by metal or electric instruments around the device, so the magnetic north is unreliable. With the camera of the device and a digital elevation model (DEM) a calibration can be carried out to determine this error.

Our preliminary work had resulted a simple edge-based algorithm for skyline extraction that does not require manual interaction. However, in some cases, this method could be improved by using artificial neural networks. In the proposed study, a feed-forward neural network trained with scaled conjugated gradient back-propagation method in Matlab has been employed to choose the presumed skyline from the candidates. This method has achieved promising results by taking advantage of typical characteristics of skylines in the specific environment.

References

- [1] Baatz, Georges, et al. Large scale visual geo-localization of images in mountainous terrain. Computer Vision ECCV 2012. Springer Berlin Heidelberg, 2012. 517-530.
- [2] Baboud, Lionel, et al. Automatic photo-to-terrain alignment for the annotation of mountain pictures. Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011.
- [3] Bazin, Jean-Charles, et al. Dynamic programming and skyline extraction in catadioptric infrared images. Robotics and Automation, 2009. ICRA'09. IEEE International Conference on. IEEE, 2009.
- [4] Behringer, Reinhold. Registration for outdoor augmented reality applications using computer vision techniques and hybrid sensors. Virtual Reality, 1999. Proceedings., IEEE. IEEE, 1999.
- [5] Boroujeni, Nasim Sepehri, S. Ali Etemad, and Anthony Whitehead. Robust horizon detection using segmentation for uav applications. Computer and Robot Vision (CRV), 2012 Ninth Conference on. IEEE, 2012.
- [6] Hung, Yao-Ling, et al. Skyline localization for mountain images. Multimedia and Expo (ICME), 2013 IEEE International Conference on. IEEE, 2013.
- [7] Yazdanpanah, Ali Pour, et al. Real-Time Horizon Line Detection based on Fusion of Classification and Clustering. International Journal of Computer Applications 121.10 (2015).

Decision-support system to maximize the robustness of computer network topologies

Teréz Nemes, Ákos Dávid, Zoltán Süle

Computer networks are supposed to meet predefined security levels. Both the definition and the measurement of such security levels are usually based on subjective methods [3]. Currently, there are no such methods based on objective parameters, which can be used to algorithmically calculate the reliability of computer networks. Reliability is often determined by the experience of IT engineers.

In this paper we propose a new method that is independent of subjective criteria. Administrators and architects can measure or improve the security level of their computer networks based on objective parameters. A decision support system has been developed by us to design computer networks not only to maximize reliability based on the current architecture, but also to satisfy financial constraints.

This paper describes the proposed methodology from the aspects of robustness. This method is based on the theory of P-graphs [1]. Former examinations show that the P-graph approach to process-network synthesis (PNS) originally conceived for conceptual design of chemical processes and supply-chain optimization provides appropriate tools for generating and analyzing structural alternatives for product supply problem. Computer network topologies are often represented by standard diagrams which do not support the robustness analysis. In order to examine the robustness of a network, a topology diagram – P-graph transformation was determined to generate the alternative structures and reliability values. Thereby, the topology given by the operator or architect can be modelled by P-graphs, and by performing optimization on this P-graph robustness can be measured [2]. The result will be reflected in the original topology, therefore we can make recommendations on the enhancement of the network's robustness. The algorithm in this paper is compared to an already published case study.

Acknowledgement

Research has been supported by the project VKSZ_12-1-2013-0088, "Development of cloud based smart IT solutions by IBM Hungary in cooperation with the University of Pannonia"

References

- [1] Botond Bertok, Karoly Kalauz, Zoltan Sule, Ferenc Friedler: Combinatorial Algorithm for Synthesizing Redundant Structures to Increase Reliability of Supply Chains: Application to Biodiesel Supply, *Industrial & Engineering Chemistry Research*, Vol. 52 Issue 1, pp. 181-186, 2013.
- [2] Ivo Raedts, Marija Petković¹, Yaroslav S. Usenko, Jan Martijn van der Werf, Jan Friso Groote, Lou Somers: Transformation of BPMN models for Behaviour Analysis, *Verification and Validation of Enterprise Information Systems*, pp. 126-137, Funchal, Madeira, Portugal 2007.
- [3] Kresimir Solic, Hrvoje Ocevcic, Marin Golub: The information systems' security level assessment model based on an ontology and evidential reasoning approach, *Computers & Security*, Vol. 55, pp. 100-112, 2015.

Active Learning using Distribution Analysis for Image Classification

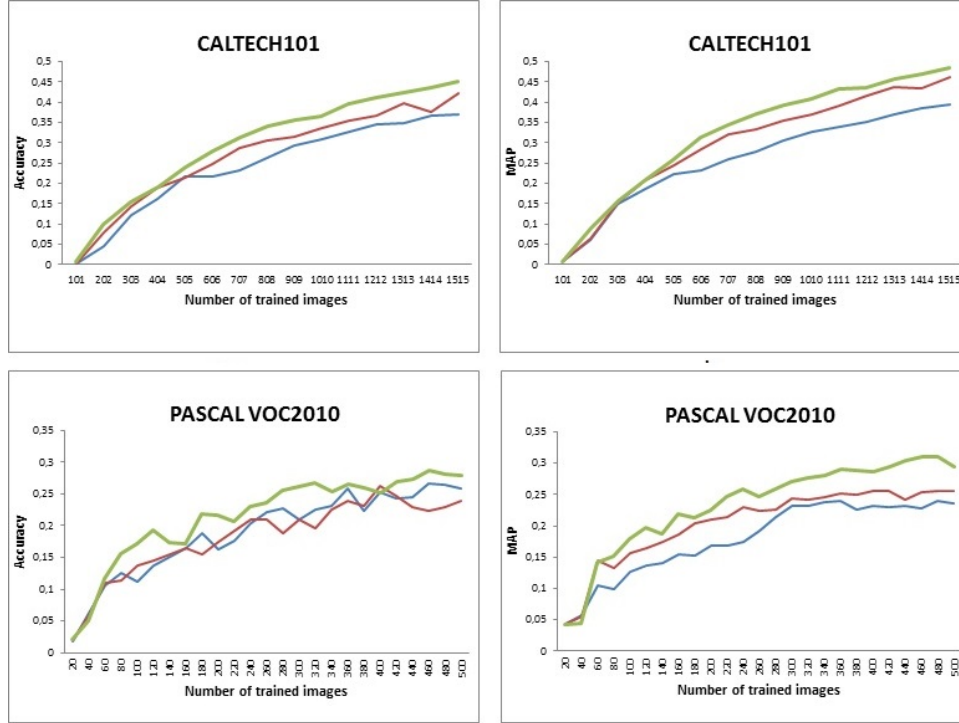
Dávid Papp, Gábor Szűcs

Nowadays, image classification is an indispensable tool for categorizing the huge amounts of visual contents both online and offline. The number of images available online increases with the rapidly growing internet usage. Besides, numerous electronic devices are capable to take a digital picture (e.g. cameras, telephones and so on), furthermore, smart devices are a click away to upload and share those pictures. This results massive data warehouses that are need to be structured, i.e. categorized. The classification of images requires labeled instances, but usually these contents are unlabeled, and labeling them is an expensive manual process. Active learning [1] is a way to address this problem since it selects a subset of the data by iteratively querying the most informative image(s) from the unlabeled ones, and then builds the classification model based on this subset instead of the whole data. In this way, the active learning algorithm aims to label as few instances as possible (i.e., minimizing the labeling cost) while it attempts to retain the same level of accuracy that could be achievable by using the total dataset. The most important question is how to estimate the informativeness of unlabeled instances, since different query strategies may lead to better or worse classification accuracy, compared to random sampling. There are many various proposed strategies in the literature, e.g. uncertainty sampling [2], query-by-committee [3], expected model change [4], expected error reduction [5]. Uncertainty sampling is a widely used, however the simplest query strategy framework, which happens to query the instances with least certainty about their labels. The focus of our work was to improve the accuracy through amplifying this technique by complementing it with a distribution analysis on the labeled dataset. For this purpose we defined a new penalty metric (see Eq. 1) which gave us an informativeness value for each unlabeled image:

$$Penalty_j = CTR_{j^*} \times \frac{1}{\#categories} \quad (1)$$

where j^* denotes the estimated category of the j^{th} image, and CTR_{j^*} increases with each iteration of queries where the received category is other then j^* . We merged these penalty values with the ones coming from uncertainty sampling, and this results the final decision scores. The advantage of this modification is a sort of balance between the classes of labeled instances. We demonstrated the efficiency of the proposed approach on two large datasets; on the PASCAL VOC2010 [6] training and validation data, and on the Caltech101 [7] image collection. We evaluated the Accuracy and the MAP (Mean Average Precision) metrics for this purpose. As can be seen in Fig. 1, our proposed approach (green line) outperforms both the general uncertainty sampling (red line) and the random sampling (blue line) methods.

Figure 1: Evaluation of the results on the Caltech101 and PASCAL VOC2010 datasets



References

- [1] Burr Settles, Active Learning Literature Survey, Computer Sciences Technical Report 1648, University of Wisconsin - Madison, 2009.
- [2] Yang, Y., Ma, Z., Nie, F., Chang, X., & Hauptmann, A. G. (2015). Multi-class active learning by uncertainty sampling with diversity maximization. *International Journal of Computer Vision*, 113(2), 113-127.
- [3] Tsai, Y. L., Tsai, R. T. H., Chueh, C. H., & Chang, S. C. (2014). Cross-Domain Opinion Word Identification with Query-By-Committee Active Learning. In *Technologies and Applications of Artificial Intelligence* (pp. 334-343). Springer International Publishing.
- [4] Cai, W., Zhang, Y., & Zhou, J. (2013, December). Maximizing expected model change for active learning in regression. In *Data Mining (ICDM), 2013 IEEE 13th International Conference on* (pp. 51-60). IEEE.
- [5] Mac Aodha, O., Campbell, N., Kautz, J., & Brostow, G. (2014). Hierarchical subquery evaluation for active learning on a graph. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 564-571).
- [6] Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J. and Zisserman, A. The PASCAL Visual Object Classes (VOC) Challenge, *International Journal of Computer Vision*, 88(2), 303-338, 2010.
- [7] L. Fei-Fei, R. Fergus and P. Perona. Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories. *IEEE. CVPR 2004, Workshop on Generative-Model Based Vision*. 2004.

Preserving Type Information in Memory Usage Measurement of C++ Programs

Zsolt Parragi, Zoltan Porkoláb

Memory profilers are essential tools to understand the dynamic behavior of complex modern programs. They help to reveal memory handling details: the wheres, the whens and the whats of memory allocations, helping programmers find memory leaks and optimizing their programs.

Most heap profilers provide sufficient information about which part of the source code is responsible for the memory allocations by showing us the relevant call stacks. The sequence of allocations inform us about their order. Heap profilers for languages with strong reflection capabilities can also show type information on the allocated objects.

C++ has no run or compile time reflection, and certain widely used constructs – for example, allocators in the standard library – hide information even more. Theoretically the information is still there, and could be recovered based on the source code, build command list and stacktrace, but this is impossible in practice.

Previous solutions used the preprocessor and operator overloading to solve this problem. Unfortunately the complex syntax of new and delete restricts the use of this method: using dynamically allocated arrays, placement new or calling the new or delete operators directly requires source code modification, even in the standard library.

In this article we will report on a type preserving heap profiler for C++ based on the LLVM/Clang tooling library which resolves the above limitations. Our technique is based on source to source compilation resulting in standard compliant C++ code. The instrumented source can be built by the same compilers as the original, without any dependency on debug symbols or run time type information.

Using this method users can tell how much memory was used of `std::vector` objects – separately for every template parameter – in every point of the run, when, where and how they were allocated and freed. Having such a type information in hand programmers can identify critical classes responsible for memory usage more easily and can perform optimizations based on evidence rather than speculations.

This information also helps in the detection of memory related bugs: for example heavy use of casts might result in calling the wrong destructor, resulting in otherwise hard to debug or detect program errors.

The tool is publicly available at <http://typegrind.github.io>

References

- [1] J. Mihalicza, Z. Porkoláb, and A. Gabor, Type-preserving heap profiler for C++ (2011)

Improving QoS in web-based distributed streaming services with applied network coding

Patrik J. Braun, Péter Ekler

With the development of the portable devices and sensors by 2020 50 billion devices [1] can be connected to the internet. This enormous increase of devices generates more and more data that has to be transmitted over the network. Conventional content distribution approaches use client-server topology, where one server or a group of server serve all the clients. The main drawbacks of this technology are poor scalability, which can lead to bad Quality of Service (QoS) in case of high client number. The servers are placed at the edge of the network, usually physically far away from the clients, which increases the latency of the network and the server is being a single point of failure. Lately a shift to Peer-to-Peer (P2P) topology can be observed. This alternative topology has several beneficial properties. A fully distributed network is self-scalable, since the more participant in the network means more data source in addition the participant are located in the network, closer to each other. Both characteristic can lower the delay and increase the throughput. In most cases creating fully distributed networks are technically not possible. A minimal central intelligence is needed to store a list of available peers in the network, else the peers cannot find each other. This central intelligence can be a single point of failure, but these servers are lightweight, easy to be replaced, and do not store any important data. The conventional web browsers use *HTTP* protocol to transfer data. The protocol is designed to be used only in client-server topology. Providers use expensive content distribution networks (CDN) for serving static content to load balance their services and lower the serve time. There have been several attempts to achieve browser-based P2P data distribution without installing any browser plugins. Among the first attempts was Adobe with Flash Player based *RTMP* protocol, which later was replaced with *RTMFP* [2] protocol. This can be considered a semi solution, since Flash Player is usually installed on the computer of the user. The first pure *JavaScript* based technology was Web Real-Time Communication (WebRTC) [3], that offers a several protocols for streaming media directly from browser to another browser. The standardization of the protocol still not finished since 2011, but Chrome, Firefox and Opera already implement it. Using these technologies for P2P content distribution, several companies and libraries were created [4] [5] [6].

In this paper we investigated the methods, how browser based P2P streaming can be achieved with the help of *WebRTC*. In our work we focus not just on data distribution, but efficient data distribution. We have applied network coding [7], specifically Random Linear Network Coding (RLNC) in the network, due to its beneficial characteristics, mainly its rate-less nature and its recode ability to create new coded data without having the whole original data. Furthermore, the coding already proved its ability in P2P environment [8] [9] [10]. We propose two protocols for efficient browser based P2P content streaming. The first protocol called *WebPeer* protocol for distributed content distribution. Based on this protocol, we design an extended protocol, called *CodedWebPeer* that support RLNC encoded packets. Both protocols make possible to download the data parallel from the server and from other peers as well. Using our protocol, a testbed is implemented to investigate the characteristic of the protocols. We have carried out several measures in different scenarios. We have investigated the impact of the network size and the clients' storage size on QoS in the network. Both parameter turned out to be important, since low storage size limits the possibilities of sharing content in a distributed network. Small network size limits the number of potential partners.

Through our results, we show that modern browsers are capable of maintaining P2P connections and carrying out complex network coding calculations. We show that employing our protocols for data streaming, more than 300% network throughput can be archived, comparing

to conventional server-client topology. We analyze the impact of the two parameters on the network throughput. We show that our *CodedWebPeer* performs at least as good as our *WebPeer* and is much less sensible for the change of the network size and the clients' storage size.

References

- [1] Dave Evans, *The Internet of Things:How the Next Evolution of the Internet is Changing Everything*, Cisco Internet Business Solutions Group (IBSG), April, 2011.
- [2] M. Thornburgh, *Adobe's Secure Real-Time Media Flow Protocol*, <http://www.rfc-editor.org/rfc/rfc7016.txt>, February, 2016
- [3] A. Bergkvist, D. C. Burnett, C. Jennings, A. Narayanan, b. Aboba, *WebRTC 1.0: Real-time Communication Between Browsers*, <https://www.w3.org/TR/webrtc/>, February, 2016.
- [4] Jie Wu,ZhiHui Lu, Bisheng Liu, Shiyong Zhang *PeerCDN: A novel P2P network assisted streaming content delivery network scheme*,2008. CIT 2008. 8th IEEE International Conference on Computer and Information Technology, Sydney, NSW, 2008
- [5] L. Zhang, F. Zhou, A. Mislove, R. Sundaram *Maygh: building a CDN from client web browsers*, EuroSys'13 Proceedings of the 8th ACM European Conference on Computer Systems,Pages 281-294, New York, NY, USA, 2013
- [6] Werner, M.J., Vogt, C., Schmidt, T.C. *Let Our Browsers Socialize: Building User-Centric Content Communities on WebRTC*, 2014 IEEE 34th International Conference on Distributed Computing Systems Workshops (ICDCSW), Madrid, Spain, June 30 2014-July 3 2014
- [7] Ahlswede, R., Ning Cai, Li, S.-Y.R., Yeung, R.W.: *Network information flow*, *Information Theory*, IEEE Transactions on (Volume:46 , Issue: 4), pages: 1204-1216,ISSN: 0018-9448, 2000
- [8] P. J. Braun M. Sipos, P. Ekler, H. Charaf, *Increasing data distribution in BitTorrent networks by using network coding techniques*, European Wireless 2015; 21th European Wireless Conference, Budapest, May, 2015.
- [9] Christos Gkantsidis, Pablo Rodriguez Rodriguez, *Network Coding for Large Scale Content Distribution* IEEE Infocom 2005
- [10] Christos Gkantsidis, John Miller, Pablo Rodriguez, *Anatomy of a P2P Content Distribution system with Network Coding*, IPTPS, 2006

Detecting Similarities in Process Data generated with Computer-based Assessment Systems

Aleksandar Pejić

Computer-based Assessment Systems are nowadays widely adopted in the field of education. They are used to evaluate, measure and document the performance of students, from as soon as the early childhood (preschool), throughout primary, secondary and higher education. A prominent example is OECD's (Organisation for Economic Co-operation and Development) Programme for International Student Assessment (PISA), an international large-scale testing program on student performance. This study is conducted worldwide every three years, and measures the performance of 15-16 years old pupils. It focuses on mathematics, reading, science and problem-solving areas of assessment. While OECD regularly publishes the results of the study across the participating countries, it also makes publicly available the databases for each year the pupils took the test.

In PISA 2012 a computer-based assessment focusing on the problem-solving skills was introduced, that is particularly interesting for researchers in applied informatics and social sciences. The CBA system recorded not only the final result of the task at hand, but all the actions that the pupils performed in the CBA environment. Thus the resulting dataset is rich with information; it makes it possible to analyze the applied problem solving strategies [1], but also sets a number of challenges for the researchers. Adequate information extracting and aggregating methods have to be identified. To obtain proper information for complex analyzes, as well as to determine its meaning in the domain of social sciences, experts from the field have to be consulted.

In this work we are investigating the process of extracting, aggregating and finding similarities between sequences of actions recorded for a PISA 2012 problem-solving item. Identifying and assembling the feature sequence for each test taker is a prerequisite for further researching and explaining cognitive performance.

One possible approach is to utilize natural language processing (NLP) methodologies. A technique for analyzing problem-solving process data based on N-grams was developed on the premise of similar structure among action sequences and word sequences in natural language [2], following the conclusion that the evolution and use of sequential models is closely related to the statistical modeling of texts [3]. Another important aspect taken from text categorization is identifying the key features, as well as disregarding the insignificant features, by applying a corresponding strategy, known as feature selection, which provides basic means for classification [4].

The described technique gives us a feature sequence for each test taker. We are following up by investigating a method for clustering the test takers based on the similarity of their sequences. In natural language processing and information retrieval a text document is often represented as a simplified multiset of its words. This model is known as bag-of-words [5], or bag-of-features in our case. After assembling a multiset of selected features, we can represent each test takers' feature sequence as a vector. These vectors are then quantified and grouped using k-means clustering. For data processing and clustering the Weka software was used (Waikato Environment for Knowledge Analysis).

References

- [1] S. Greiff, S. WÄzstenberg, F. Avvisati. Computer-generated log-file analyses as a window into students'minds? A showcase study based on the PISA 2012 assessment of problem solving. Elsevier Computers & Education, 91, 92-105, 2015

- [2] Q. He, M. von Davier. Identifying Feature Sequences from Process Data in Problem-Solving Items with N-Grams. Springer Proceedings in Mathematics & Statistics, 140, 179-196, 2014
- [3] G. A. Fink. Markov models for pattern recognition. Berlin, Germany, Springer, 2008
- [4] S. Li, R. Xia, C. Zong & C. Huang. A framework of feature selection methods for text categorization. In Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP, 692-700, 2009
- [5] G. Salton, M.J. McGill. Introduction to modern information retrieval. McGraw-Hill, 1983

Taxonomy and Survey of IoT Cloud Use Cases

Tamás Pflanzner, Attila Kertész

Cloud computing [1] enables flexible resource provisions that has become hugely popular for many businesses to take advantage of responding quickly to new demands from customers. Internet of Things (IoT) systems recently appeared as a dynamic global network infrastructure with self configuring capabilities [2], in which things can interact and communicate among themselves and with the environment through the Internet by exchanging sensor data, and react autonomously to events and influence them by triggering actions with or without direct human intervention. According to recent reports in this area (e.g. [3]), there will be 30 billion devices always online and more than 200 billion devices discontinuously online by 2020. Such estimations call for an ecosystem that provides means to interconnect and control these devices with cloud solutions, thus user data can be stored in a remote location and can be accessed from anywhere. There is a growing number of cloud providers offering IoT-specific services, since cloud computing has the potential to satisfy IoT needs such as hiding data generation, processing and visualization tasks. While each provider offers its own set of features, two critical features they all have in common are the ability to connect devices and to store the data generated by those devices.

In this paper we study and investigate 9 IoT cloud providers in detail, and identify 4 use cases arose from joint utilization of IoT and cloud systems. The related works suggest that the following IoT cloud cases can be derived: 1 – a local, ad-hoc IoT system that can be formed from near-by things (e.g. smart watch) to perform a certain task. 2 – a cooperative IoT Cloud system, in which some tasks of an application run in a smart device, and some run in the cloud. 3 – a bridged IoT cloud system, where a smart device can act as a bridge or gateway to collect and move sensor data to the cloud. Finally, 4 – a direct IoT cloud system, where a thing (such as a smart TV) communicates with the cloud directly.

IoT application developers do not only have to decide which cloud provider to use, but they also have to choose which combination of protocols and data structures best fits their application. To aid the design, development and testing processes of these systems, an IoT device simulator could be useful to emulate real devices or sensors. Therefore we also gather the requirements for basic functionalities of such a simulator, i.e. to send and receive messages, generate sensor data (for one or more devices), and react to received messages.

The main contributions of this work are: (i) a survey and classification of IoT cloud providers, (ii) a taxonomy of IoT cloud use cases represented by cloud-assisted IoT applications and (iii) a proposal for an IoT device simulator to support further research on IoT device management. Our future work will address the development of a generic IoT device simulator to aid IoT application development with the surveyed providers, based on the identified use cases and requirements. Such a simulator can be used in the future to help cloud application developers to learn IoT device handling and to evaluate test IoT environments without buying real sensors.

References

- [1] Buyya B, Yeo C S, Venugopal S, Broberg J, and Brandic I, Cloud computing and emerging it platforms: Vision, hype, and reality for delivering computing as the 5th utility, *Future Generation Computer Systems*, vol. 25, no. 6, pp. 599-616, June 2009.
- [2] H. Sundmaeker, P. Guillemin, P. Friess, S. Woelffle. Vision and challenges for realising the Internet of Things. CERP IoT report. CN: KK-31-10-323-EN-C, March 2010.
- [3] J. Mahoney and H. LeHong, The Internet of Things is coming, Gartner report. Online: <https://www.gartner.com/doc/1799626/internet-things-coming>, September 2011.

Long-acting insulin management for blood glucose prediction model

Rebaz Ahmed Hama Karim

Introduction: Reliable methods for short term blood glucose level prediction can be efficiently integrated into mobile Ambient Assisted Living (AAL) services for diabetes management. Subcutaneous insulin absorption is one of the key factors beside the meal consumed and the blood glucose history values that such prediction methods must consider. According to the current practice, insulin dependent diabetic patients use a basal or long-acting insulin injection once a day that produces a steady insulin level for the whole day, and bolus injections for every meal in order to control their blood glucose. The mathematical models commonly used blood glucose prediction can handle only bolus insulin in a correct way, which makes effective only for inpatient care. One can simulate the effect of the basal insulin as a single injection of bolus insulin with a slowly rising curve with a very high maximum value. However, in reality the basal insulin level reaches a proper level in a very short time and the effect lasts for a long time. An adjustment to the model is required to properly handle basal insulin in order to reach more accurate results. By substituting the single big dose of basal insulin with a series of smaller bolus insulin doses we can carry out the steady curve of insulin presence in the blood similar to the curve defined by the medicine manufacturers.

Method: We implemented our blood glucose prediction model [1] by combining two state-of-the-art models. The first one models the glucose absorption in two compartments (stomach and intestines) [2]. This model can manage various sorts of food with different glycemic indices and it can also handle the overlap of meals. The other model is based on nonlinear discrete-delay differential equations [3], and it model the blood glucose and insulin control system. The equations of the model describe insulin transfer between subcutaneous insulin depots, insulin absorption into blood and the role of insulin in blood glucose control. In our proposed method, we model the long-acting insulin in a correct way by using a series of smaller insulin doses, instead of one big dose, the original time interval divided into short subintervals with low maximal absorption time (T_{max}). We tested the method on a data set of a clinical trial in which 16 insulin dependent patients (7 female and 9 male) used a Continuous Glucose Monitor (CGM) device to record their blood glucose levels for six days, while their meals were recorded. In order to test the proposed method, we used 4 profiles of basal insulins, each profile having a specific number of doses, dose amounts, injection times and T_{max} values, thus simulating the original basal insulin injection. We applied the original model and the corrected model to predict short term (180 mins) blood glucose levels with various basal insulin profiles and computed the average absolute error of the prediction compared to the CGM records.

Results: The average improvement achieved by the correction was 0.63 mmol/l and the biggest improvement was 2.37 mmol/l. In some special cases the average error increased due to the improved basal insulin model (in case of Lantus and Levemir insulins). Table I show the results.

	<u>Profile 1</u>	<u>Profile 2</u>	<u>Profile 3</u>	<u>Profile 4</u>
<u>Average improvement (mmol/l)</u>	0.57	0.63	0.44	0.55
<u>Biggest improvement (mmol/l)</u>	2.37	2.35	2.32	2.35
<u>Worse results (data set)</u>	3	4	5	3

TABLE I. Average improvement, Biggest improvement and Worse results data for each insulin profile

We also analyzed the prediction accuracy at the next meal time. A frequently used measure in the literature is the ratio of error, i.e. the rate of prediction errors less than 3 mmol/l. We

can state that in 5 cases more than 50% of the errors were within 3 mmol/l, in 11 cases it was between 30% and 50%. In two cases, it failed to achieve an acceptable result. Figure I shows results of two day's test for P11 patient.



FIGURE I. Prediction of Blood Glucose Level at the next meal time for P11 patient.

Conclusion: It is difficult to compare our results for long-acting insulin management in blood glucose prediction with the literature because to our best knowledge no research has been published in exactly this field so far.

Our results without correction were not far from the best results published [4]. The average improvement of more than 0.6 mmol/l earned by the correction is remarkable. This improvement, combined with the expected improvement from the planned management of other factors like stress, insulin sensitivity and physical activity, could make our prediction model more efficient and reliable as a module of the Lavinia lifestyle mirror mobile application [5] for outpatient healthcare.

References

- [1] P. Gyuk, T. Lorincz, Rebaz A. H. K., I. Vassanyi. Diabetes Lifestyle Support with Improved Glycemia Prediction Algorithm. Telemedicine, and Social Medicine (eTELEMED 2015), ISBN 978-1-61208-384-1, Lisboa, Portugal, 22-27 February 2015, pp. 95-100.
- [2] T. Arleth, S. Andreassen, M. Orsini-Federici, A. Timi, and M. Massi-Benedetti, Optimisation of a model of glucose absorption from mixed meals, 2005. [http://www.mmds.aau.dk/Publicationer/DIAS/optimisation of a model of glucose absorption from mixed meals.pdf](http://www.mmds.aau.dk/Publicationer/DIAS/optimisation%20of%20a%20model%20of%20glucose%20absorption%20from%20mixed%20meals.pdf)
- [3] P. Palumbo, P. Pepe, S. Panunzi, and A. De Gaetano. Glucose control by subcutaneous insulin administration: a DDE modelling approach, in Proc. 18th IFAC World Congress, Milan, Italy, 2011, pp. 1471-1476
- [4] R. A. H. Karim, P. Gyuk, T. Lorincz, I. Vassanyi, I. Kosa. Blood Glucose Level Prediction for Mobile Lifestyle Counseling. Med-e-Tel2015, Int. Conf. of the International Society for Telemedicine and eHealth, 22-24 April 2015, Luxembourg. Knowledge Resources, Vol 8 (2015), ISSN 1998-5509, pp. 8-12.
- [5] I. Kosa, I. Vassanyi, M. Nemes, K.H. Kalmanne, B. Pinter, and L. Kohut, A fast, android based dietary logging application to support the life change of cardio-metabolic patients, Global Telemedicine and eHealth Updates: Knowledge Resources, 7, 2014, pp. 553-556 <http://www.lavinia.hu/>.

Trends and paradigm change in ERP datacenter management

Selmeci Attila, Orosz Tamás Gábor

In today's world each of the companies use ERP (Enterprise Resource Planning) systems, except the smallest micro companies. In a retail, utility, manufacturing or other bigger company (or even institute) the ERP is not the only system, but they have at least some core business solutions, warehousing system or some components of the ERP II model (SRM, CRM, SCM, PLM, etc.), and a business analytic solution. These systems are communicating with each other. Some of them are front systems communication with end-users, or customers, others only do the daily work in the background. These systems should be managed, controlled and maintained correctly. Some companies build their own data centers, others place their equipment into external datacenters, some of them outsource their application management as well, and the newest ones use virtual environments, services, infrastructures from the cloud. Each of the possible datacenters should be well designed, well organized, and well managed. The management frequently stops at the edge of the hardware infrastructure. It means only monitoring, alerting, some time information about system / application failure or stop. The business requires a bit more; it wants good performance, always-available services, flexible infrastructure on logical and physical level as well, improved IT costs, etc.

The ERP (Enterprise Resource Planning) systems (and generally the others as well) as most of the bigger applications are built according to client-server architecture. The software-oriented view of the architecture generally connects three parts of the system: database layer, application layer and presentation layer. According to newer ERP or business application the application layer is mainly a web server with special content, or portal server. The older applications, like SAP have a huge engine with its own interface to GUI (Graphical User Interface). The presentation layer for the older application has thick client or sometimes web based thin client as well, but the newer ones are more filigree by implementing the newest technologies. All of them are using database layer, but the technique and content may differ. The hardware-oriented view of ERP systems' client-server architecture shows us different level of infrastructures. In each one we can meet network elements and active components, storage layer and computing elements (servers, virtual machines, zones, etc.). In this architecture not only one-one connections are available, but one-many as well. As an example on the software-oriented view we can have one database and many applications servers, or many clients connecting to one application server. On the hardware-oriented view an example for 1-n connection is the physical host, which can be mounted into a rack or blade center, or one storage area offered via NFS (Network File System) can be attached to several operating systems. Many-many connections can be also found in software layer if we would handle parallel queries, e.g. using Oracle RAC (Real Application Cluster) with multiple clients. The software and hardware-oriented layers are not strictly separated, because the modern hardware elements can be controlled, managed by software tools. These tools offer more and more features for the given hardware elements. A well-known example can be a hypervisor based virtualization environment, where many (n) physical hosts can run many (m) virtual machines.

Each part of this high level conglomerate can be handled as an object. We should think of them as composite object, because they are not simple ones, but contain many smaller, important parts (just like a storage contains the single HDDs, SSDs, PCI Express caches, or even internal operating system, etc.). The objects have special relationships via their connections to each other. All these lead us to think about an object oriented model of a datacenter. The model describes the infrastructure element and the software elements as well containing their properties and statuses as well. The model is the first step to the new management, because it only describes the datacenter, the used infrastructure and the software, applications and services running on the infrastructure.

Our goal is to build a management tool using the object model to flexible maintain the software defined elements, monitor the environment, alert on request and proactively offer tasks or even actions to be executed automatically. The object model enables these proactive tasks by having events and handlers for them. A very simple example can be well during the night a huge data load is started into the data warehouse system without previously notifying the operators. The operating system based host agent sends data to the object, which checks the predefined thresholds. If there is not enough place is available an event is raised to notify the storage object to allocate more place to the used volume. Of course the management tool should notify the operators as well about the emergency volume resize function. In the first level we try to monitor, control and manage services like database, application instance. We implement and adaptive solution, which runs services elastically distributed on physical or virtual hosts in the environment. Parallel we would refine, refactor the object model to converge our abstraction to the real datacenter architecture, elements and their links to each other. Our future plan is to collect and analyze IT data, because the great power in such a solution is not only the monitoring and automation, but also the analytic and the behavior changes according the analysis.

Validation of Well-formedness Constraints on Uncertain Models

Oszkár Semeráth, Dániel Varró

In Model Driven Development (MDD) models are the main design artifacts, from which documentation, system configuration, or even source code can be automatically generated. As a result, MDD is widely used in industry in various domains including business modeling, avionics or automotive. However, the development of large models is still a challenging task, as multiple design rules have to be satisfied simultaneously. Additionally, several design decisions have to be made in an early stage of the model development in order to create valid models, even if the developer is uncertain about which possibility is the most favorable (e.g. which class should contain a particular attribute). This problem makes the development of initial prototype models difficult, and hides possible valid design options in complex models.

The goal of our research is to create new validation technique which supports the development of uncertain models by checking the well-formedness constraints. Existing partial modeling techniques like [1] allows a modeler to explicitly express model uncertainty, or concretize possible design candidate [2], but does not support well-formedness constraints. On the other hand, there is efficient tool support for defining and checking well-formedness constraints and design rules over (fully specified) model instances using graph pattern matching [3]. Our technique combines these two approach by transforming graph queries of the target modeling language to graph queries for partial models, in order to match on malformed, possibly malformed or correct model partitions. Our technique uses the partial snapshot formalism of [4], the pattern language of VIATRA [3], and is compatible with EMF [5], which is the de facto modeling standard in MDD.

Therefore, design rules specified for concrete models can be automatically checked for uncertain models to (i) detect invalid elements, (ii) filter invalid design options. Additionally, (iii) model generation techniques [4, 2] can be supported by efficiently evaluating several predicates with a query engine.

References

- [1] Famelis, Michalis and Salay, Rick and Chechik, Marsha, *Partial Models: Towards Modeling and Reasoning with Uncertainty*, Proceedings of the 34th International Conference on Software Engineering, pp. 573-583, 2012.
- [2] Salay, Rick and Famelis, Michalis and Chechik, Marsha, *Language Independent Refinement Using Partial Modeling*, Fundamental Approaches to Software Engineering, pp. 224-239, 2012.
- [3] Gábor Bergmann, Zoltán Ujhelyi, István Ráth, Dániel Varró, *A Graph Query Language for EMF Models*, Theory and Practice of Model Transformations, ICMT 2011, pp. 167-182, 2011.
- [4] Semeráth, Oszkár, Barta, Ágnes, Horváth, Ákos, Szatmári, Zoltán, Varró, Dániel, *Formal validation of domain-specific languages with derived features and well-formedness constraints*, Software & Systems Modeling, pp. 1-36, 2015.
- [5] The Eclipse Project, *Eclipse Modeling Framework*, <http://www.eclipse.org/emf>

On the Efficiency of Text-based Comparison and Merging of Software Models

Ferenc Attila Somogyi

The importance of model-driven engineering in software development has been steadily increasing in the last few years, as models are more often used in the industry for various purposes. Models can be described and edited via either a graphical or a textual interface. The graphical approach usually makes the model more easily readable, but writing efficiency does not scale well with model complexity. It is the more convenient and common approach to use, especially if the model has to be interpreted by third-party users as well. On the other hand, the textual approach is rarely used compared to the graphical one. It can be harder to read at first (especially for third-party users), but writing efficiency scales well with model complexity. Generally speaking, using a textual representation to describe and edit the model pays off when the model is substantial in size or complexity, especially if there are efficient development tools available for the editing process. The availability of such tools results in the editing process being more similar to traditional source code-based development. During source code-based development, teamwork is supported by version control systems that help manage the different versions of source code files. The same idea can be applied to models and their textual representations in order to facilitate teamwork during model-driven engineering. In previous work [1], we briefly presented a method that can be used to compare and merge the textual representations of software models. We also compared this method to other, already existing model comparing or merging approaches. In addition to the raw texts, the method uses the abstract syntax trees (AST-s) built from the texts as basis of the comparison. This makes the comparison a lot more accurate, as AST-s give more accurate information than the raw texts. The textual representations are also needed though, as discovering some conflicts require textual comparison in addition to comparing the AST-s. The method is not dependent on any particular modeling environment or formal language used to describe the textual representations. This is achieved by demanding certain operations from the parser of the language. Thus, the method can be the foundation of conflict handling in version control systems that support the textual representations of models. The main difference between the presented method and other, specialized model comparing and merging approaches (like EMF Compare) is generality, meaning that most approaches only work with a specific modeling environment. Another difference is the fact that the presented method aims to compare the textual representations instead of the structure of the model. This results in more overall accuracy while still maintaining generality (with certain restrictions). In this paper, we examine the different steps of the method in detail with the focus being on efficiency and performance. These are important considerations if the method is to be used in real version control systems. We also consider and review alternatives to different steps of the method by examining their advantages and disadvantages.

References

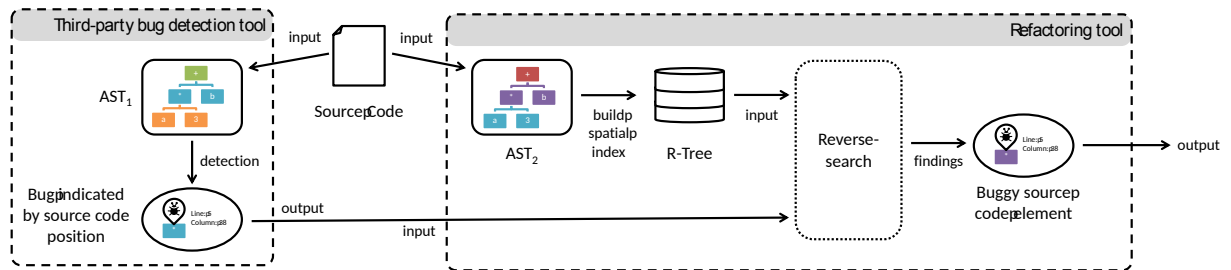
- [1] F.A. Somogyi. Merging Textual Representations of Software Models, *MultiScience - 2016. microCAD International Multidisciplinary Scientific Conference*, under publish.
- [2] EMF Compare. <https://www.eclipse.org/emf/compare/>.

Finding Matching Source Code Elements in an AST Using Position Information

Gábor Szőke

To decrease software maintenance cost software development companies use static source code analysis techniques. Static analysis tools are capable of finding potential bugs, anti-patterns, coding rule violations and they can enforce coding style standards. To achieve this the tools create a hierarchical representation of the source code called *Abstract Syntax Tree* (AST). In this structure each node of the tree denotes a construct occurring in the source code. The AST can be represented in various ways and most static analysis tools build their own AST representation, typically which fits them the best for their given purpose.

Sometimes it is necessary to create a mapping between two different ASTs, e.g. finding matching source code elements in an AST using just position information. We met this problem in a project where we developed a refactoring tool which was meant to be able to refactor coding issues based on source code position. In this study, we present an approach which addresses this problem. Our solution takes source code position and type information of a node from one AST and uses this information to *reverse search* the matching node on the other AST. To make reverse searching possible, a *spatial database* is used, which is created by transforming the source code into *geometric space*. Line numbers and column positions from the AST are used to define areas. These areas are used to create *R-trees*, where area based reverse searching is possible.



To evaluate our approach we use the output of the *PMD* [1] source code analyzer tool and use its bug report's position information as the source data and then we pick the *Columbus* [2] AST as target AST to find the matching source code element in the syntax tree. We used this technique in a project, where the found element was used as input in an automated refactoring transformation on the target AST. The transformation modified the syntax tree in a way which fixed the bug reported by the *PMD* static analyzer tool. As the final step, the refactored version of the source code got generated from the target AST and developers could review the changes.

The evaluation showed that our approach can be adapted to a real-life scenario, and it can provide viable results. Our tool was used in a project where it assisted in more than 4,000 automated refactorings covering systems ranging from 200 to 2,500 kLOC.

References

- [1] PMD website: <https://pmd.github.io/>
- [2] Ferenc, Rudolf and Beszédes, Árpád and Tarkiainen, Mikko and Gyimóthy, Tibor, *Columbus – Reverse Engineering Tool and Schema for C++*, Proceedings of the 18th International Conference on Software Maintenance (ICSM'02), pp. 172-181, 2002.

Classification based Symbolic Indoor Positioning over the Miskolc IIS Dataset

Judit Tamás, Zsolt Tóth

Performance analysis of classification based symbolic indoor positioning methods are presented in this paper. Symbolic positioning can be considered as a classification task, where the classes are the positions, and the attributes are the measured values. The ILONA (Indoor Localization and Navigation) system was used to record a hybrid indoor positioning dataset that was used for the evaluation of the tested methods. The goal of this performance analysis is to determine the most accurate classification based symbolic indoor positioning method for the ILONA system. The results can be used to improve the currently used positioning algorithm of the ILONA system.

Hybrid indoor positioning systems simultaneously use various sensors and technologies [1]. Hybrid systems were designed to overcome the limitation of the standard indoor positioning systems that are based on a single technology, such as Bluetooth [2, 3], GSM [4], WLAN [5], Magnetometer [6], RFID [7], infrared [8] and ultrasonic [9]. Hybrid systems can be based on the fingerprinting [5] approach, which is popular among indoor positioning systems [10]. Data mining algorithms are often used in fingerprinting based solutions. The usage of multiple sensor data could increase the accuracy of the used data mining algorithm. Hybrid systems can achieve higher performance than the standard solutions because they use multiple technologies.

Miskolc IIS Hybrid Indoor Positioning System Dataset [11] was used to evaluate the selected indoor positioning methods. The dataset contains more than 1500 measurements about a 3-story building. Bluetooth, WiFi RSSI and Magnetometer sensors were used to create the dataset. Absolute and symbolic position is given for each measurement in the dataset. Symbolic position can be considered as a class label, so the positioning task can be converted into a classification problem. This paper focuses on the performance of the standard classification methods, such as k -NN, Naive Bayes, Decision Tree and Rule induction. Rapid Miner was used to evaluate the performance of these methods for symbolic indoor positioning. This dataset allows the evaluation and comparison of various indoor positioning methods.

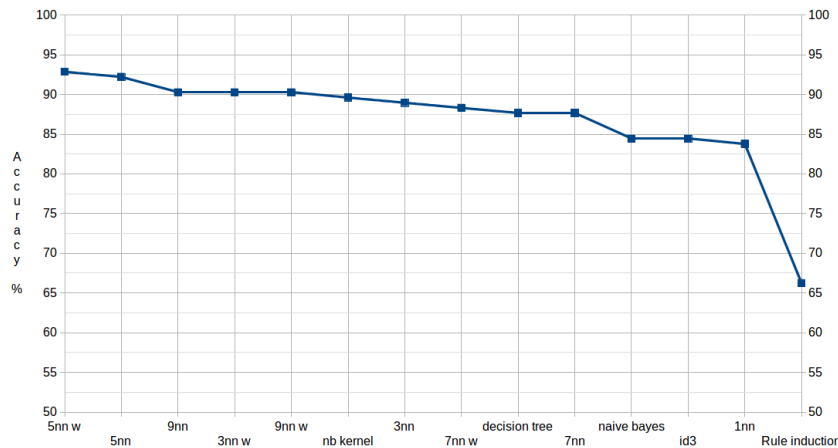


Figure 1: Performances of the tested methods

Experimental results show that the k -NN classifiers are superior than the other tested methods as seen in Figure 1. The $kNN w$ denotes the k -NN algorithm with weighted vote based on the distance. The *nb kernel* stands for the Naive Bayes kernel algorithm, where each

attribute considered as individual attribute. The highest achieved accuracy was 93% by the weighted k -NN algorithm, where the k parameter was 5. The performance of the k -NN algorithm was depended on the k parameter. The k -NN algorithm had the lowest performance when the k parameter was chosen to 1. The usage of kernel function in the case of Naive Bayes classifier has increased the accuracy with approximately 5%. The performance of the decision tree depended on the building algorithm. The Rule induction achieved the lowest performance in this test. The gradient of the lines denotes the degree of the accuracy decay. Based on the experimental results, the implementation of the k -NN and weighted k -NN algorithms in the ILONA system is suggested.

References

- [1] Hui Liu, Houshang Darabi, Pat Banerjee, and Jing Liu. Survey of wireless indoor positioning techniques and systems. *Systems, Man, and Cybernetics, Part C: Applications and Reviews*, IEEE Transactions on, 37(6):1067–1080, 2007.
- [2] Anja Bakkeliën, Michel Deriaz, and Stephane Marchand-Maillet. Bluetooth indoor positioning, Master’s thesis, University of Geneva, 2012.
- [3] Jakub Neburka, Zdenek Tlamsa, Vlastimil Benes, Ladislav Polak, Ondrej Kaller, Libor Bolecek, Jiri Sebesta and Tomas Kratochvil. Study of the Performance of RSSI based Bluetooth Smart Indoor Positioning, *RADIOELEKTRONIKA 2016 26th International Conference*.
- [4] SS Wang, Marilyn Green, and M Malkawa. E-911 location standards and location commercial services. In *Emerging Technologies Symposium: Broadband, Wireless Internet Access*, 2000 IEEE, pages 5–pp. IEEE, 2000.
- [5] Moustafa Youssef and Ashok Agrawala. The horus wlan location determination system. In *Proceedings of the 3rd international conference on Mobile systems, applications, and services*, ACM, 2005.
- [6] Simo Särkkä, Ville Tolvanen, Juho Kannala, and Esa Rahtu. Adaptive kalman filtering and smoothing for gravitation tracking in mobile systems. 2005.
- [7] Lionel M Ni, Yunhao Liu, Yiu Cho Lau, and Abhishek P Patil. Landmarc: indoor location sensing using active rfid. *Wireless networks*, 10(6):701–710, 2004.
- [8] Roy Want and Andy Hopper. Active badges and personal interactive computing objects. *Consumer Electronics*, IEEE Transactions on, 38(1):10–20, 1992.
- [9] Ultrasonic based positioning. <http://www.hexamite.com/hx11.htm>.
- [10] Hakan Koyuncu and Shuang Hua Yang. A survey of indoor positioning and object locating systems. *IJCSNS International Journal of Computer Science and Network Security*, 2010.
- [11] Zsolt Tóth, Judit Tamás. Miskolc IIS Hybrid IPS: Dataset for Hybrid Indoor Positioning , *RADIOELEKTRONIKA 2016 26th International Conference*.

Regression Model Building and Efficiency Prediction of Header Compression Implementations for VoIP

Máté Tömösközi

Modern cellular networks utilising the long-term evolution (LTE) and the coming 5G set of standards face an ever-increasing demand for low-latency mobile data from connected devices. Header compression is employed to minimise the overhead for IP-based cellular network traffic, thereby decreasing the overall bandwidth usage and, subsequently, transmission delays.

Albeit both version of Robust Header Compression achieve around 80–90 % gain (see [1] and [2]), neither of these compression designs account for the varying channel conditions that can be encountered in wireless setups (e.g., as a result of weak signals or interference). Specifically, the RFC 5525 for RoHCv2 states that “A compressor always [...] repeats the same type of update until it is fairly confident that the decompressor has successfully received the information.” and “The number of repetitions that is needed to obtain this confidence is normally related to the packet loss and out-of-order delivery characteristics of the link where header compression is used; it is thus not defined in this document [RFC 5525].” The impact of these repetitions is closely related to the compressed IP-streams’ protocol field dynamics and will differ from application to application.

Our current research efforts try to address this issue by enabling current compressor implementations to configure themselves online, thereby making the compression adaptable to changing channel conditions and various network stream characteristics, which will in turn minimise compression overhead and will pave the way for the adaptation of header compression in IoT scenarios. Our previous efforts target the preliminary analysis of compression utilities and their prediction with linear regression in [3], which has resulted in limited prediction accuracy.

We evaluate various prediction models employing R^2 and error scores next to complexity (number of coefficients) based on an RTP specific training data set and a separately captured live VoIP audio call. We find that the proposed weighted Ridge regression model explains about 70 % of the training data and 72 % of a separate VoIP transmission’s utility. This approach outperforms the Ridge and first-order Bayesian regressions by up to 50 % and the second and third order regressions utilising polynomial basis functions by up to 20 %, making it well-suited for utility estimation.

References

- [1] M. Tomoskozi, P. Seeling, F. H. Fitzek, *Performance evaluation and comparison of RObust Header Compression (ROHC) ROHCv1 and ROHCv2 for multimedia delivery*, Workshops Proceedings of the Global Communications Conference, GLOBECOM, pp. 1346-1350, 2013.
- [2] M. Tömösközi and P. Seeling and P. Ekler and F. H. P. Fitzek, *Performance Evaluation and Implementation of IP and Robust Header Compression Schemes for TCP and UDP Traffic in the Wireless Context*, Engineering of Computer Based Systems (ECBS-EERC), pp. 45-50, 2015.
- [3] Máté Tömösközi and Patrick Seeling and Péter Ekler and Frank H.P. Fitzek, *Performance Prediction of Robust Header Compression version 2 for RTP Audio Streaming Using Linear Regression*, European Wireless 2016 (EW2016), 2016.

Stable Angina Clinical Pathway Correlation Clustering

Zsolt Vassy, István Kósa PhD, MD, István Vassányi PhD

Introduction

The aim of the current study was to apply statistical and network science techniques [2] to depict how the clinical pathways of patients with chest pain affect their survival. Our study relies on GYEMSZI's data [1], from the years between 2004 and 2008. We included the data of 506.087 patients who underwent diagnostic procedures related to ischemic heart disease. The patients were assigned to one of the 136 *de facto* primary health care centers based on their residence.

We distinguished three different investigative processes: non-imaging cardiac stress test i.e. stress ECG (E), non-invasive imaging methods (NI) and invasive imaging methods (I). The clinical pathways are built up from the combination of these three processes. The traditional pathway is E-NI-I, but the doctors have the freedom to skip the E/NI steps for patients with higher coronary artery disease risk or inability to perform the noninvasive imaging or non-imaging tests.

Data overview

In the vast majority of cases, non-imaging cardiac stress test (E) was the only procedure performed. The ratio of non-invasive and invasive imaging methods increases together with the age of patients.

Cluster analysis

Network building

The primary care centers were compared with each other using Pearson's correlation according to the distribution of different clinical pathways. We have made a network based on a correlation matrix in which nodes are primary care centers and edge weights are correlation coefficients. We calculated all of the 18769 correlation coefficients with a 95 % confidence level.

Clustering

Using Louvain clustering (a modularity-based clustering) method on this network, 3 different care center groups were identified.

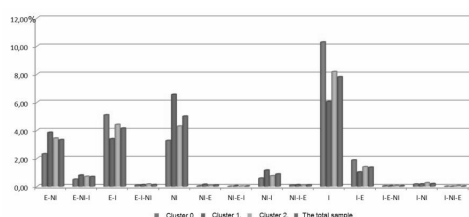


Figure 1: Clinical pathway spectrum of clusters (clinical pathway: x-axis, percentage of occurrence: y-axis)

The 3 clusters have different characteristics (Fig 1.):

- Cluster 0: relative preference for invasive imaging
- Cluster 1: relative preference for non-invasive procedures
- Cluster 2: relative preference for invasive treatment followed by non-invasive imaging

We calculated the revascularization and mortality rates for all of the clusters. Cluster 0 (the "invasive" cluster) has a much higher revascularization rate than Cluster 1 ($p < 0.01$), but the 365-day mortality rates for the two clusters are almost the same ($p < 0.05$), see Fig 1. This indicates that in many cases the revascularization procedure may be unnecessary.

We have observed a correlation between the spatial position of health care centers and the cluster membership (see Fig 2.). Cluster 0 was dominant in Western Hungary, Cluster 1 in Eastern Hungary and Cluster 2 in Central Hungary.

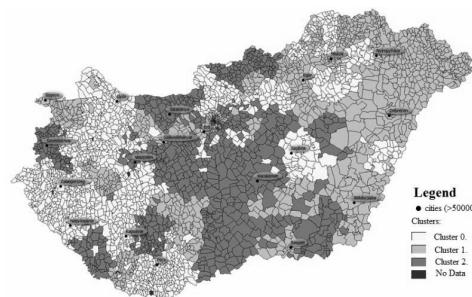


Figure 2: Position of health care centers belonging to the three clusters

Discussion and conclusions

Previous studies analyzed only the health care profiles of individual care centers or even single cases of patients [1]. In our study, the large number of cases allows the clustering and cluster-wise analysis of the profiles in a statistically meaningful way, thus higher level conclusions on the efficiency and organization of the health care system can be drawn. The correlation between the spatial position of health care centers and cluster membership suggests that there is a kind of information spread between neighboring institutions.

There are considerable differences in the utilization of patient evaluation pathways through the country, which is connected with similar differences in the subsequent revascularization procedures. In spite of this difference, 365-day mortality rates do not differ significantly between Cluster 0 and Cluster 1. It would be worthwhile to review revascularization practices for primary health care centers of Cluster 0.

References

- [1] I.Kósa, A.Nemes, É.Belicza, F.Király, I.Vassányi. Regional differences in the utilisation of coronary angiography as initial investigation for the evaluation of patients with suspected coronary artery disease. *Int J Cardiol* 2013; 168: 5012-5. DOI: <http://dx.doi.org/10.1016/j.ijcard.2013.07.148>
- [2] Tariq Ahmad, Nihar Desai et al., Clinical Implications of Cluster Analysis-Based Classification of Acute Decompensated Heart Failure and Correlation with Bedside Hemodynamic Profiles. *PLOS ONE* February 3, 2016

Parallel implementation of a modular, population based global optimizer package

Dániel Zombori, Balázs Bánhelyi

In the world of optimization the problems are very different. Therefore it is impossible to give a single solution method to find the global optimum for all of them. The best way is to create a package of various optimization methods, which contains solvers for a wide range of problems. The other issue is that sometimes much CPU time is necessary to compute the objective function in a point. For example, when the objective values come from simulations, which is very common in physics problems. In this case the running time of the optimization method is very high. A suitable solution for this is to compute the function evaluations simultaneously. In this case the algorithm have to handle the objective values simultaneously. Hence the trivial simple deterministic methods is not fitting for this problem.

In our presentation we parallelize the GLOBAL optimization method which is developed by in the Institute of Informatics, University of Szeged. GLOBAL is a stochastic technique that is a sophisticated composition of sampling, clustering, and local search.

The talk presents the architecture of the stochastic global optimization algorithm GLOBAL and the single thread Java implementation. Then we present the parallel clustering method which applied in our solution. After this, we show the sampling, clustering, and local search methods working in parallel. We applied a technique which is based on the priority of the earlier mentioned methods, and all threads determine alone the next methods to complete.

Finally we illustrate the efficiency of our method on large scale popular test functions.

LIST OF AUTHORS

Rebaz, Ahmed Hama Karim: University of Pannonia, Hungary,
E-mail: rebazah@gmail.com

Babati, Bence: Eötvös Loránd University, Hungary,
E-mail: babati@caesar.elte.hu

Babau, Cristian-Nicolae: Politehnica University of Timisoara, Romania,
E-mail: cristian.babau@student.upt.ro

Bagóczki, Zsolt: University of Szeged, Hungary,
E-mail: Bagoczki.Zsolt@stud.u-szeged.hu

Baráth, Áron: Eötvös Loránd University, Hungary,
E-mail: baratharon@caesar.elte.hu

Baumgartner, János: University of Pannonia, Hungary,
E-mail: baumgartner@dc.s.uni-pannon.hu

Bogdándy, Bence: University of Miskolc, Hungary,
E-mail: Bence.bogdandy@hotmail.hu

Braun, Patrik: Budapest University of Technology and Economics, Hungary,
E-mail: braun.patrik@aut.bme.hu

Brunner, Tibor: Eötvös Loránd University, Hungary,
E-mail: bruntib@caesar.elte.hu

Budai, Ádam: Budapest University of Technology and Economics, Hungary,
E-mail: budai8adam@gmail.com

Bán, Dénes: University of Szeged, Hungary,
E-mail: zealot@inf.u-szeged.hu

Búr, Márton: Budapest University of Technology and Economics, Hungary,
E-mail: bur@mit.bme.hu

Fekete, Tamás: Budapest University of Technology and Economics, Hungary,
E-mail: feketet@aut.bme.hu

Fábián, Gábor: Eötvös Loránd University, Hungary,
E-mail: robagnaibaf@gmail.com

Garai, Ábel: University of Debrecen, Hungary,
E-mail: garai.abel@inf.unideb.hu

Gazdi, László: Budapest University of Technology and Economics, Hungary,
E-mail: gazdilaci@gmail.com

Gyimesi, Péter: University of Szeged, Hungary,
E-mail: pgyimesi@inf.u-szeged.hu

Györffy, Lajos: University of Szeged, Hungary,
E-mail: gyorffylajos38@gmail.com

Gál, Péter: University of Szeged, Hungary,
E-mail: galpeter@inf.u-szeged.hu

Hajdu, László: University of Szeged, Hungary,
E-mail: hajdul@jgypk.u-szeged.hu

Horváth, Ferenc: University of Szeged, Hungary,
E-mail: hferenc@inf.u-szeged.hu

Kabódi, László: Budapest University of Technology and Economics, Hungary,
E-mail: kabodil@gmail.com

Kalmár, György: University of Szeged, Hungary,
E-mail: kalmargy@inf.u-szeged.hu

Katona, Melinda: University of Szeged, Hungary,
E-mail: mkatona@inf.u-szeged.hu

Koós, Krisztián: University of Szeged, Hungary,
E-mail: koos.krisztian@brc.mta.hu

Kádár, István: University of Szeged, Hungary,
E-mail: ikadar@inf.u-szeged.hu

Lipták, Levente: Budapest University of Technology and Economics, Hungary,
E-mail: liptaklevente@gmail.com

Mester, Abigél: University of Szeged, Hungary,
E-mail: mester.abigel@gmail.com

Mishra, Biswajeeban: University Of Szeged , Hungary,
E-mail: mishra@inf.u-szeged.hu

Molnár, Csaba: Biological Research Centre, Hungary,
E-mail: molnar.csaba@mta.brc.hu

Márton, Gábor: Eötvös Loránd University, Hungary,
E-mail: martongabesz@gmail.com

Nagy, Balázs: Corvinus University of Budapest, Hungary,
E-mail: balazs.nagy@outlook.com

Nemes, Teréz: Budapest Business School, Hungary,
E-mail: nemes.terez@uni-bge.hu

Papp, David: Budapest University of Technology and Economics, Hungary,
E-mail: pappd@tmit.bme.hu

Parragi, Zsolt: Eötvös Loránd University, Hungary,
E-mail: zsoltparragi@caesar.elte.hu

Pejic, Aleksandar: University of Szeged, Hungary,
E-mail: pejic@inf.u-szeged.hu

Pflanzner, Tamás: University of Szeged, Hungary,
E-mail: tampfla@inf.u-szeged.hu

Pomázi, Krisztián: Budapest University of Technology and Economics, Hungary,
E-mail: krisztianpomazi@gmail.com

Selmeci, Attila: Óbuda University, Hungary,
E-mail: selmeci.attila@amk.uni-obuda.hu

Semeráth, Oszkár: Budapest University of Technology and Economics, Hungary,
E-mail: semerath@mit.bme.hu

Somogyi, Ferenc Attila: Budapest University of Technology and Economics, Hungary,
E-mail: ferencsomogyi92@gmail.com

Szőke, Gábor: University of Szeged, Hungary,
E-mail: kancsuki@inf.u-szeged.hu

Szűcs, Judit: University of Szeged, Hungary,
E-mail: szucsjudit92@gmail.com

Tamás, Judit: University of Miskolc, Hungary,
E-mail: judit0310@gmail.com

Tóth, László: University of Szeged, Hungary,
E-mail: tothl@jgypk.u-szeged.hu

Tóth, Zsolt: University of Miskolc, Hungary,
E-mail: tothzs@iit.uni-miskolc.hu

Tömösközi, Máté: Budapest University of Technology and Economics, Hungary,
E-mail: tomoskozi.mate@aut.bme.hu

Vassy, Zsolt: University of Pannonia, Hungary,
E-mail: zsolt.vassy@gmail.com

Zombori, Dániel: University of Szeged, Hungary,
E-mail: zombdaniel@gmail.com

NOTES

CS²